OFFICE OF
Educational Technology

Artificial Intelligence and the Future of Teaching and Learning

Insights and Recommendations

May 2023

Artificial Intelligence and the Future of Teaching and Learning

Miguel A. Cardona, Ed.D.
Secretary, U.S. Department of Education

Roberto J. Rodríguez
Assistant Secretary, Office of Planning, Evaluation, and Policy Development

Kristina Ishmael
Deputy Director, Office of Educational Technology

May 2023

How to Cite
While permission to reprint this publication is not necessary, the suggested citation is as follows:
U.S. Department of Education, Office of Educational Technology, *Artificial Intelligence and Future of Teaching and Learning: Insights and Recommendations*, Washington, DC, 2023.

This report is available at https://tech.ed.gov

Table of Contents

Introduction

The U.S. Department of Education (Department) is committed to supporting the use of technology to improve teaching and learning and to support innovation throughout educational systems. This report addresses the clear need for sharing knowledge and developing policies for "Artificial Intelligence," a rapidly advancing class of foundational capabilities which are increasingly embedded in all types of educational technology systems and are also available to the public. We will consider "educational technology" (edtech) to include both (a) technologies specifically designed for educational use, as well as (b) general technologies that are widely used in educational settings. Recommendations in this report seek to engage teachers, educational leaders, policy makers, researchers, and educational technology innovators and providers as they work together on pressing policy issues that arise as Artificial Intelligence (AI) is used in education.

AI can be defined as "automation based on associations." When computers automate reasoning based on associations in data (or associations deduced from expert knowledge), two shifts fundamental to AI occur and shift computing beyond conventional edtech: (1) from capturing data to *detecting patterns* in data and (2) from providing access to instructional resources to *automating decisions* about instruction and other educational processes. Detecting patterns and automating decisions are leaps in the level of responsibilities that can be delegated to a computer system. The process of developing an AI system may lead to bias in how patterns are detected and unfairness in how decisions are automated. Thus, educational systems must govern their use of AI systems. This report describes opportunities for using AI to improve education, recognizes challenges that will arise, and develops recommendations to guide further policy development.

Rising Interest in AI in Education

Today, many priorities for improvements to teaching and learning are unmet. Educators seek technology-enhanced approaches addressing these priorities that would be safe, effective, and scalable. Naturally, educators wonder if the rapid advances in technology in everyday lives could help. Like all of us, educators use AI-powered services in their everyday lives, such as voice assistants in their homes; tools that can correct grammar, complete sentences, and write essays; and automated trip planning on their phones. Many educators are actively exploring AI tools as they are newly released to the public[1]. Educators see opportunities to use AI-powered capabilities like speech recognition to increase the support available to students with disabilities, multilingual learners, and others who could benefit from greater adaptivity and personalization in digital tools for learning. They are exploring how AI can enable writing or improving lessons, as well as their process for finding, choosing, and adapting material for use in their lessons.

Educators are also aware of new risks. Useful, powerful functionality can also be accompanied with new data privacy and security risks. Educators recognize that AI can automatically produce output that is inappropriate or wrong. They are wary that the associations or automations created by AI may amplify unwanted biases. They have noted new ways in which students may

[1] Walton Family Foundation (March 1, 2023). Teachers and students embrace ChatGPT for education. https://www.waltonfamilyfoundation.org/learning/teachers-and-students-embrace-chatgpt-for-education

represent others' work as their own. They are well-aware of "teachable moments" and pedagogical strategies that a human teacher can address but are undetected or misunderstood by AI models. They worry whether recommendations suggested by an algorithm would be fair. Educators' concerns are manifold. Everyone in education has a responsibility to harness the good to serve educational priorities while also protecting against the dangers that may arise as a result of AI being integrated in edtech.

To develop guidance for edtech, the Department works closely with educational constituents. These constituents include educational leaders—teachers, faculty, support staff, and other educators—researchers; policymakers; advocates and funders; technology developers; community members and organizations; and, above all, learners and their families/caregivers. Recently, through its activities with constituents, the Department noticed a sharp rise in interest and concern about AI. For example, a 2021 field scan found that developers of all kinds of technology systems—for student information, classroom instruction, school logistics, parent-teacher communication, and more—expect to add AI capabilities to their systems. Through a series of four listening sessions conducted in June and August 2022 and attended by more than 700 attendees, it became clear that constituents believe that action is required now in order to get ahead of the expected increase of AI in education technology—and they want to roll up their sleeves and start working together. In late 2022 and early 2023, the public became aware of new generative AI chatbots and began to explore how AI could be used to write essays, create lesson plans, produce images, create personalized assignments for students, and more. From public expression in social media, at conferences, and in news media, the Department learned more about risks and benefits of AI-enabled chatbots. And yet this report will not focus on a specific AI tool, service, or announcement, because AI-enabled systems evolve rapidly. Finally, the Department engaged the educational policy expertise available internally and in its relationships with AI policy experts to shape the findings and recommendations in this report.

Three Reasons to Address AI in Education Now

> *"I strongly believe in the need for stakeholders to understand the cyclical effects of AI and education. By understanding how different activities accrue, we have the ability to support virtuous cycles. Otherwise, we will likely allow vicious cycles to perpetuate."*
> *—Lydia Liu*

During the listening sessions, constituents articulated three reasons to address AI now:

First, AI may enable achieving educational priorities in better ways, at scale, and with lower costs. Addressing varied unfinished learning of students due to the pandemic is a policy priority, and AI may improve the adaptivity of learning resources to students' strengths and needs. Improving teaching jobs is a priority, and via automated assistants or other tools, AI may provide teachers greater support. AI may also enable teachers to extend the support they offer to individual students when they run out of time. Developing resources that are responsive to the knowledge and experiences students bring to their learning—their community and cultural assets—is a priority, and AI may enable greater customizability of curricular resources to meet local needs.

As seen in voice assistants, mapping tools, shopping recommendations, essay-writing capabilities, and other familiar applications, AI may enhance educational services.

Second, urgency and importance arise through awareness of system-level risks and anxiety about potential future risks. For example, students may become subject to greater surveillance. Some teachers worry that they may be replaced—to the contrary, the Department firmly rejects the idea that AI could replace teachers. Examples of discrimination from algorithmic bias are on the public's mind, such as a voice recognition system that doesn't work as well with regional dialects, or an exam monitoring system that may unfairly identify some groups of students for disciplinary action. Some uses of AI may be infrastructural and invisible, which creates concerns about transparency and trust. AI often arrives in new applications with the aura of magic, but educators and procurement policies require that edtech show efficacy. AI may provide information that appears authentic, but actually is inaccurate or lacking a basis in reality. Of the highest importance, AI brings new risks in addition to the well-known data privacy and data security risks, such as the risk of scaling pattern detectors and automations that result in "algorithmic discrimination" (e.g., systematic unfairness in the learning opportunities or resources recommended to some populations of students).

Third, urgency arises because of the scale of possible unintended or unexpected consequences. When AI enables instructional decisions to be automated at scale, educators may discover unwanted consequences. In a simple example, if AI adapts by speeding curricular pace for some students and by slowing the pace for other students (based on incomplete data, poor theories, or biased assumptions about learning), achievement gaps could widen. In some cases, the quality of available data may produce unexpected results. For example, an AI-enabled teacher hiring system might be assumed to be more objective than human-based résumé scoring. Yet, if the AI system relies on poor quality historical data, it might de-prioritize candidates who could bring both diversity and talent to a school's teaching workforce.

In summary, it is imperative to address AI in education *now* to realize key opportunities, prevent and mitigate emergent risks, and tackle unintended consequences.

Toward Policies for AI in Education

The 2023 AI Index Report from the Stanford Institute for Human-Centered AI has documented notable acceleration of investment in AI as well as an increase of research on ethics, including issues of fairness and transparency.[2] Of course, research on topics like ethics is increasing because problems are observed. Ethical problems will occur in education, too.[3] The report found a striking interest in 25 countries in the number of legislative proposals that specifically include AI. In the United States, multiple executive orders are focused on ensuring AI is trustworthy and equitable, and the White House Office of Science and Technology Policy has introduced a

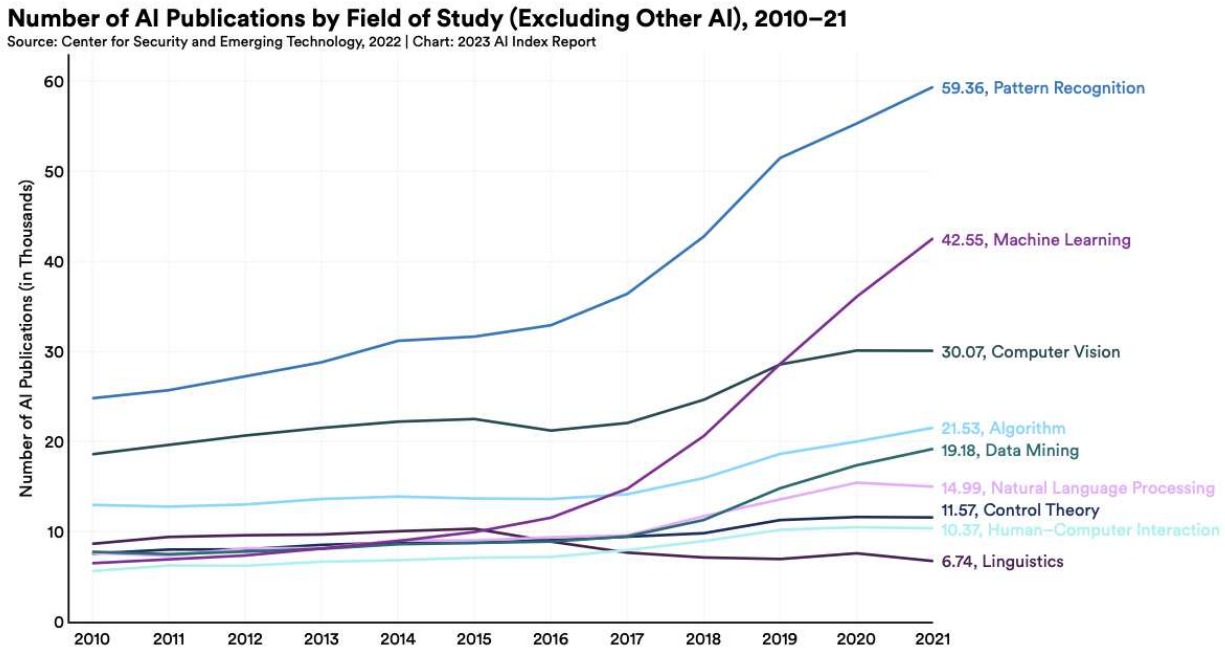
[2] Maslej, N., Fattorini, L., Brynjolfsson E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J.C., Parli, V., Shoham, Y., Wald, R., Clark, J. and Perrault, R., (2023). *The AI index 2023 annual report.* Stanford University: AI Index Steering Committee, Institute for Human-Centered AI.

[3] Holmes, W. & Porayska-Pomsta, K. (Eds.) (2022). *The ethics of artificial intelligence in education.* Routledge. ISBN 978-0367349721

Blueprint for an AI Bill of Rights (Blueprint)[4] that provides principles and practices that help achieve this goal. These initiatives, along with other AI-related policy activities occurring in both the executive and legislative branches, will guide the use of AI throughout all sectors of society. In Europe, the European Commission recently released *Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators*.[5]

AI is moving fast and heralding societal changes that require a national policy response. In addition to broad policies for all sectors of society, education-specific policies are needed to address new opportunities and challenges within existing frameworks that take into consideration federal student privacy laws (such as the Family Educational Rights and Privacy Act, or FERPA), as well as similar state related laws. AI also makes recommendations and takes actions automatically in support of student learning, and thus educators will need to consider how such recommendations and actions can comply with laws such as the Individuals with Disabilities Education Act (IDEA). We discuss specific policies in the concluding section.

Figure 1: Research about AI is growing rapidly. Other indicators, such as dollars invested and number of people employed, show similar trends.



AI is advancing exponentially (see Figure 1), with powerful new AI features for generating images and text becoming available to the public, and leading to changes in how people create text and

[4] White House Office of Science and Technology Policy (October 2022), *Blueprint for an AI bill of rights: Making automated systems work for the American people.* The White House Office of Science and Technology Policy. https://www.whitehouse.gov/ostp/ai-bill-of-rights/

[5] European Commission, Directorate-General for Education, Youth, Sport and Culture. (2022). *Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators*, Publications Office of the European Union. **https://data.europa.eu/doi/10.2766/153756**

images[6]. The advances in AI are not only happening in research labs but also are making news in mainstream media and in educational-specific publications.

Researchers have articulated a range of concepts and frameworks for ethical AI[7], as well as for related concepts such as equitable, responsible, and human-centered AI. Listening session participants called for building on these concepts and frameworks but also recognized the need to do more; participants noted a pressing need for guardrails and guidelines that make educational use of AI advances safe, especially given this accelerating pace of incorporation of AI into mainstream technologies. As policy development takes time, policy makers and educational constituents together need to start now to specify the requirements, disclosures, regulations, and other structures that can shape a positive and safe future for all constituents—especially students and teachers.

Policies are urgently needed to implement the following:

1. leverage automation to advance learning outcomes while protecting human decision making and judgment;

2. interrogate the underlying data quality in AI models to ensure fair and unbiased pattern recognition and decision making in educational applications, based on accurate information appropriate to the pedagogical situation;

3. enable examination of how particular AI technologies, as part of larger edtech or educational systems, may increase or undermine equity for students; and

4. take steps to safeguard and advance equity, including providing for human checks and balances and limiting any AI systems and tools that undermine equity.

[6] Sharples, M. & Pérez y Pérez, R. (2022). *Story machines: How computers have become creative writers*. Routledge. ISBN 9780367751951

[7] Akgun, S., Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI Ethics,* 2, 431–440. https://doi.org/10.1007/s43681-021-00096-7

Building Ethical, Equitable Policies Together

In this report, we aim to build on the listening sessions the Department hosted to engage and inform all constituents involved in making educational decisions so they can prepare for and make better decisions about the role of AI in teaching and learning. AI is a complex and broad topic, and we are not able to cover everything nor resolve issues that still require more constituent engagement. This report is intended to be a starting point.

The opportunities and issues of AI in education are equally important in K-12, higher education, and workforce learning. Due to scope limitations, the examples in this report will focus on K-12 education. The implications are similar at all levels of education, and the Department intends further activities in 2023 to engage constituents beyond K-12 schools.

Guiding Questions

Understanding that AI increases automation and allows machines to do some tasks that only people did in the past leads us to a pair of bold, overarching questions:

1. What is our collective vision of a desirable and achievable educational system that leverages automation to advance learning while protecting and centering human agency?
2. How and on what timeline will we be ready with necessary guidelines and guardrails, as well as convincing evidence of positive impacts, so that constituents can ethically and equitably implement this vision widely?

In the Learning, Teaching, and Assessment sections of this report, we elaborate on elements of an educational vision grounded in what today's learners, teachers, and educational systems need, and we describe key insights and next steps required. Below, we articulate four key foundations for framing these themes. These foundations arise from what we know about the effective use of educational technology to improve opportunity, equity, and outcomes for students and also relate to the new Blueprint.

Foundation 1: Center People (Parents, Educators, and Students)

Education-focused AI policies at the federal, state, and district levels will be needed to guide and empower local and individual decisions about which technologies to adopt and use in schools and classrooms. Consider what is happening in everyday lives. Many of us use AI-enabled products because they are often better and more convenient. For example, few people want to use paper maps anymore; people find that technology helps us plan the best route to a destination more efficiently and conveniently. And yet, people often do not realize how much privacy they are giving up when they accept AI-enabled systems into their lives. AI will bring privacy and other risks that are hard to address only via individual decision making; additional protections will be needed.

There should be clear limits on the ability to collect, use, transfer, and maintain our personal data, including limits on targeted advertising. These limits should put the burden on platforms to minimize how much information they collect, rather than burdening Americans with reading fine print.[8]

As protections are developed, we recommend that policies center people, not machines. To this end, a first recommendation in this document (in the next section) is an emphasis on **AI with humans in the loop**. Teachers, learners, and others need to retain their agency to decide what patterns mean and to choose courses of action. The idea of humans in the loop builds on the concept of "Human Alternatives, Consideration, and Fallback" in the Blueprint and ethical concepts used more broadly in evaluating AI, such as preserving human dignity. A top policy priority must be establishing human in the loop as a requirement in educational applications, despite contrary pressures to use AI as an alternative to human decision making. Policies should not hinder innovation and improvement, nor should they be burdensome to implement. Society needs an education-focused AI policy that protects civil rights and promotes democratic values in the building, deployment, and governance of automated systems to be used across the many decentralized levels of the American educational system.

Foundation 2: Advance Equity

"AI brings educational technology to an inflection point. We can either increase disparities or shrink them, depending on what we do now."
—Dr. Russell Shilling

A recent Executive Order[9] issued by President Biden sought to strengthen the connection among racial equity, education and AI, stating that "members of underserved communities—many of whom have endured generations of discrimination and disinvestment—still confront significant barriers to realizing the full promise of our great Nation, and the Federal Government has a responsibility to remove these barriers" and that the Federal Government shall both "pursue educational equity so that our Nation's schools put every student on a path to success" and also "root out bias in the design and use of new technologies, such as artificial intelligence." A specific vision of equity, such as described in the Department's recent report, *Advancing Digital Equity for All*[10] is essential to policy discussion about AI in education. This report defines digital equity as

---

[8] The White House (September 8, 2022). Readout of White House listening session on tech platform accountability. https://www.whitehouse.gov/briefing-room/statements-releases/2022/09/08/readout-of-white-house-listening-session-on-tech-platform-accountability/

[9] The White House (February 17, 2023). Executive order on further advancing racial equity and support for underserved communities through the federal government. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/02/16/executive-order-on-further-advancing-racial-equity

[10] U.S. Department of Education, Office of Educational Technology (2022). Advancing digital equity for all: Community-based recommendations for developing effective digital equity plans to close the digital divide and enable technology-empowered learning. US Department of Education.

"the condition in which individuals and communities have the information technology capacity that is needed for full participation in the society and economy of the United States."

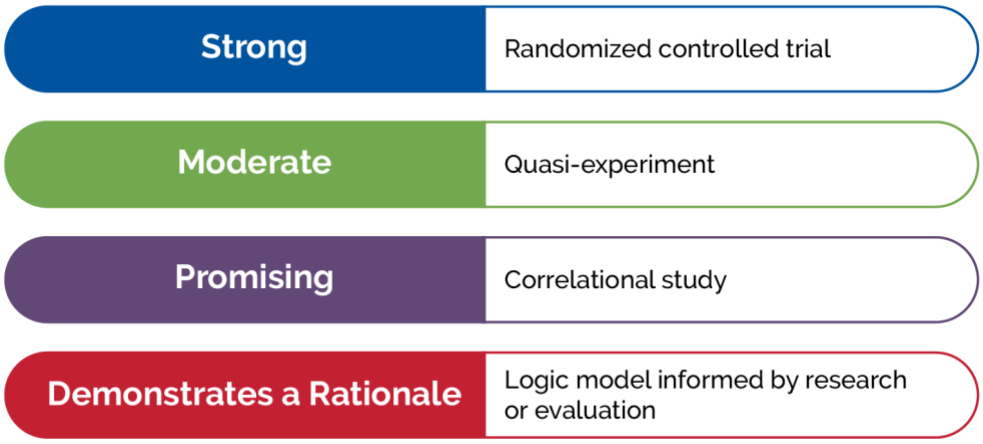
Issues related to racial equity and unfair bias were at the heart of every listening session we held. In particular, we heard a conversation that was increasingly attuned to issues of data quality and the consequences of using poor or inappropriate data in AI systems for education. Datasets are used to develop AI, and when they are non-representative or contain undesired associations or patterns, resulting AI models may act unfairly in how they detect patterns or automate decisions. Systematic, unwanted unfairness in how a computer detects patterns or automates decisions is called "algorithmic bias." Algorithmic bias could diminish equity at scale with unintended discrimination. As this document discussed in the *Formative Assessment* section, this is not a new conversation. For decades, constituents have rightly probed whether assessments are unbiased and fair. Just as with assessments, whether an AI model exhibits algorithmic bias or is judged to be fair and trustworthy is critical as local school leaders make adoption decisions about using AI to achieve their equity goals.

We highlight the concept of "algorithmic discrimination" in the Blueprint. Bias is intrinsic to how AI algorithms are developed using historical data, and it can be difficult to anticipate all impacts of biased data and algorithms during system design. The Department holds that biases in AI algorithms must be addressed when they introduce or sustain unjust discriminatory practices in education. For example, in postsecondary education, algorithms that make enrollment decisions, identify students for early intervention, or flag possible student cheating on exams must be interrogated for evidence of unfair discriminatory bias—and not only when systems are designed, but also later, as systems become widely used.

Foundation 3: Ensure Safety, Ethics, and Effectiveness

A central safety argument in the Department's policies is the need for data privacy and security in the systems used by teachers, students, and others in educational institutions. The development and deployment of AI requires access to detailed data. This data goes beyond conventional student records (roster and gradebook information) to detailed information about what students do as they learn with technology and what teachers do as they use technology to teach. AI's dependence on data requires renewed and strengthened attention to data privacy, security, and governance (as also indicated in the Blueprint). As AI models are not generally developed in consideration of educational usage or student privacy, the educational application of these models may not be aligned with the educational institution's efforts to comply with federal student privacy laws, such as FERPA, or state privacy laws.

Figure 2: The Elementary and Secondary Education Act defines four levels of evidence.



Further, educational leaders are committed to basing their decisions about the adoption of educational technology on evidence of effectiveness—a central foundation of the Department's policy. For example, the requirement to base decisions on evidence also arises in the Elementary and Secondary Education Act (ESEA), as amended, which introduced four tiers of evidence (see Figure 2). Our nation's research agencies, including the Institute of Education Sciences, are essential to producing the needed evidence. The Blueprint calls for evidence of effectiveness, but the education sector is ahead of that game: we need to insist that AI-enhanced edtech rises to meet ESEA standards as well.

Foundation 4: Promote Transparency

The central role of complex AI models in a technology's detection of patterns and implementation of automation is an important way in which AI-enabled applications, products, and services will be different from conventional edtech. The Blueprint introduces the need for transparency about AI models in terms of disclosure ("notice") and explanation. In education, decision makers will need more than notice—they will need to understand how AI models work in a range of general educational use cases, so they can better anticipate limitations, problems, and risks.

AI models in edtech will be approximations of reality and, thus, constituents can always ask these questions: **How precise are the AI models? Do they accurately capture what is most important? How well do the recommendations made by an AI model fit educational goals? What are the broader implications of using AI models at scale in educational processes?**

Building on what was heard from constituents, the sections of this report develop the theme of evaluating the quality of AI systems and tools using multiple dimensions as follows:

- **About AI**: AI systems and tools must respect data privacy and security. Humans must be in the loop.
- **Learning**: AI systems and tools must align to our collective vision for high-quality learning, including equity.
- **Teaching**: AI systems and tools must be inspectable, explainable, and provide human alternatives to AI-based suggestions; educators will need support to exercise professional judgment and override AI models, when necessary.

- **Formative Assessment**: AI systems and tools must minimize bias, promote fairness, and avoid additional testing time and burden for students and teachers.
- **Research and Development**: AI systems and tools must account for the context of teaching and learning and must work well in educational practice, given variability in students, teachers, and settings.
- **Recommendations**: Use of AI systems and tools must be safe and effective for students. They must include algorithmic discrimination protections, protect data privacy, provide notice and explanation, and provide a recourse to humans when problems arise. The people most affected by the use of AI in education must be part of the development of the AI model, system, or tool, even if this slows the pace of adoption.

We return to the idea that these considerations fit together in a comprehensive perspective on the quality of AI models in the *Recommendations* section.

Overview of Document

We begin in the next section by elaborating a definition of AI, followed by addressing learning, teaching, assessment, and research and development. Organizing key insights by these topics keeps us focused on exploring implications for improving educational opportunity and outcomes for students throughout the report.

Within these topics, three important themes are explored:

1. **Opportunities and Risks.** Policies should focus on the most valuable educational advances while mitigating risks.

2. **Trust and Trustworthiness.** Trust and safeguarding are particularly important in education because we have an obligation to keep students out of harm's way and safeguard their learning experiences.

3. **Quality of AI Models.** The process of developing and then applying a model is at the heart of any AI system. Policies need to support evaluation of the qualities of AI models and their alignment to goals for teaching and learning during the processes of educational adoption and use.

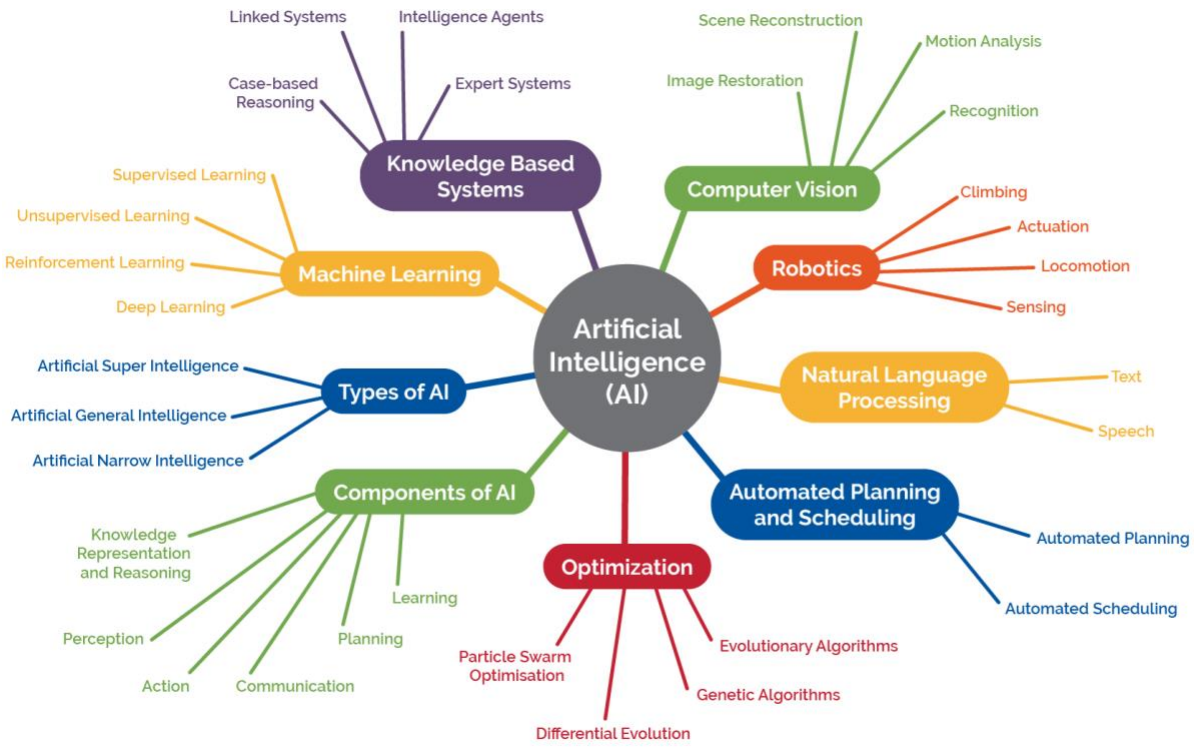
"AI in education can only grow at the speed of trust."
—Dr. Dale Allen

What is AI?

Our preliminary definition of AI as automation based on associations requires elaboration. Below we address three additional perspectives on what constitutes AI. Educators will find these different perspectives arise in the marketing of AI functionality and are important to understand when evaluating edtech systems that incorporate AI. One useful glossary of AI for Education terms is the CIRCLS Glossary of Artificial Intelligence Terms for Educators.[11]

AI is not one thing but an umbrella term for a growing set of modeling capabilities, as visualized in Figure 3.

Figure 3: Components, types, and subfields of AI based on Regona et al (2022).[12]



[11] Search for "AI Glossary Educators" to find other useful definitions.

[12] Regona, Massimo & Yigitcanlar, Tan & Xia, Bo & Li, R.Y.M. (2022). Opportunities and adoption challenges of AI in the construction industry: A PRISMA review. *Journal of Open Innovation Technology Market and Complexity, 8(45).* https://doi.org/10.3390/joitmc8010045

Perspective: Human-Like Reasoning

> *"The theory and development of computer systems able to perform tasks normally requiring human intelligence such as, visual perception, speech recognition, learning, decision-making, and natural language processing."* [13]

Broad cultural awareness of AI may be traced to the landmark 1968 film "2001: A Space Odyssey"—in which the "*H*euristically-programmed *AL*gorithmic" computer, or "HAL," converses with astronaut Frank. HAL helps Frank pilot the journey through space, a job that Frank could not do on his own. However, Frank eventually goes outside the spacecraft, HAL takes over control, and this does not end well for Frank. HAL exhibits human-like behaviors, such as reasoning, talking, and acting. Like all applications of AI, HAL can help humans but also introduces unanticipated risks—especially since AI reasons in different ways and with different limitations than people do.

The idea of "human-like" is helpful because it can be a shorthand for the idea that computers now have capabilities that are very different from the capabilities of early edtech applications. Educational applications will be able to converse with students and teachers, co-pilot how activities unfold in classrooms, and take actions that impact students and teachers more broadly. There will be both opportunities to do things much better than we do today and risks that must be anticipated and addressed.

The "human-like" shorthand is not always useful, however, because AI processes information differently from how people process information. When we gloss over the differences between people and computers, we may frame policies for AI in education that miss the mark.

Perspective: An Algorithm that Pursues a Goal

> *"Any computational method that is made to act independently towards a goal based on inferences from theory or patterns in data."* [14]

This second definition emphasizes that AI systems and tools identify patterns and choose actions to achieve a given goal. These pattern recognition capabilities and automated recommendations will be used in ways that impact the educational process, including student learning and teacher instructional decision making. For example, today's personalized learning systems may recognize signs that a student is struggling and may recommend an alternative instructional sequence. The scope of pattern recognition and automated recommendations will expand.

---

[13] IEEE-USA Board of Directors. (February 10, 2017). *Artificial intelligence research, development and regulation.* IEEE http://globalpolicy.ieee.org/wp-content/uploads/2017/10/IEEE17003.pdf

[14] Friedman, L., Blair Black, N., Walker, E., & Roschelle, J. (November 8, 2021) *Safe AI in education needs you*. Association of Computing Machinery blog, https://cacm.acm.org/blogs/blog-cacm/256657-safe-ai-in-education-needs-you/fulltext

Correspondingly, humans must determine the types and degree of responsibility we will grant to technology within educational processes, which is not a new dilemma.

For decades, the lines between the role of teachers and computers have been discussed in education, for example, in debates using terms such as "'computer-aided instruction," "blended instruction," and "personalized learning." Yet, how are instructional choices made in systems that include both humans and algorithms? Today, AI systems and tools are already enabling the adaptation of instructional sequences to student needs to give students feedback and hints, for example, during mathematics problem solving or foreign language learning. This discussion about the use of AI in classroom pedagogy and student learning will be renewed and intensify as AI-enabled systems and tools advance in capability and become more ubiquitous.

Let's start with another simple example. When a teacher says, "Display a map of ancient Greece on the classroom screen," an AI system may choose among hundreds of maps by noting the lesson objectives, what has worked well in similar classrooms, or which maps have desirable features for student learning. In this case, when an AI system suggests an instructional resource or provides a choice among a few options, the instructor may save time and may focus on more important goals. However, there are also forms of AI-enabled automation that the classroom instructor may reject, for example, enabling an AI system or tool to select the most appropriate and relevant readings for students associated with a historical event. In this case, an educator may choose not to utilize AI-enabled systems or tools given the risk of AI creating false facts ("hallucinating") or steering students toward inaccurate depictions of historical events found on the internet. Educators will be weighing benefits and risks like these daily.

Computers process theory and data differently than humans. AI's success depends on associations or relationships found in the data provided to an algorithm during the AI model development process. Although some associations may be useful, others may be biased or inappropriate. Finding bad associations in data is a major risk, possibly leading to algorithmic discrimination. Every guardian is familiar with the problem: A person or computer may say, "Our data suggests your student should be placed in this class," and the guardian may well argue, "No, you are using the wrong data. I know my child better, and they should instead be placed in another class." This problem is not limited exclusively to AI systems and tools, but the use of AI models can amplify the problem when a computer uses data to make a recommendation because it may appear to be more objective and authoritative, even if it is not.

Although this perspective can be useful, it can be misleading. A human view of agency, pursuing goals, and reasoning includes our human abilities to make sense of multiple contexts. For example, a teacher may see three students each make the same mathematical error but recognize that one student has an Individualized Education Program to address vision issues, another misunderstands a mathematical concept, and a third just experienced a frustrating interaction on the playground; the same instructional decision is therefore not appropriate. However, AI systems often lack data and judgement to appropriately include context as they detect patterns and automate decisions. Further, case studies show that technology has the potential to quickly derail from safe to unsafe or from effective to ineffective when the context shifts even slightly. For this and other reasons, people must be involved in goal setting, pattern analysis, and decision-making.[15]

---

[15] Russell, S. (2019). Human compatible: Artificial intelligence and the problem of control. Viking. ISBN 978-0-525-55861-3.

Perspective: Intelligence Augmentation

> *"Augmented intelligence is a design pattern for a human-centered partnership model of people and artificial intelligence (AI) working together to enhance cognitive performance, including learning, decision making, and new experiences."* [16]

Foundation #1 (above) keeps humans in the loop and positions AI systems and tools to support human reasoning. "Intelligence Augmentation" (IA)[17] centers "intelligence" and "decision making" in humans but recognizes that people sometimes are overburdened and benefit from assistive tools. AI may help teachers make better decisions because computers notice patterns that teachers can miss. For example, when a teacher and student agree that the student needs reminders, an AI system may provide reminders in whatever form a student likes without adding to the teacher's workload. Intelligence Automation (IA) uses the same basic capabilities of AI, employing associations in data to notice patterns, and, through automation, takes actions based on those patterns. However, IA squarely focuses on helping people in human activities of teaching and learning, whereas AI tends to focus attention on what computers can do.

Definition of "Model"

The above perspectives open a door to making sense of AI. Yet, to assess AI meaningfully, constituents must consider specific models and how they are developed. In everyday usage, the term "model" has multiple definitions. We clarify our intended meaning, which is a meaning similar to "mathematical model," below. (Conversely, note that "model" as used in "AI model" is unlike the usage in "model school" or "instructional model" as AI model is not a singular case created by experts to serve as an exemplar.)

AI models are like financial models: an approximation of reality that is useful for identifying patterns, making predictions, or analyzing alternative decisions. In a typical middle school math curriculum, students use a mathematical model to analyze which of two cell phone plans is better. Financial planners use this type of model to provide guidance on a retirement portfolio. At its heart, AI is a highly advanced mathematical toolkit for building and using models. Indeed, in well-known chatbots, complex essays are written one word at a time. The underlying AI model predicts which next words would likely follow the text written so far; AI chatbots use a very large statistical model to add one likely word at a time, thereby writing surprisingly coherent essays.

When we ask about the model at the heart of AI, we begin to get answers about "what aspects of reality does the model approximate well?" and "how appropriate is it to the decision to be made?" One could similarly ask about algorithms—the specific decision-making processes that an AI model uses to go from inputs to outputs. One could also ask about the quality of the data used to build the model—for example, how representative is that data? Switching among three terms—

---

[16] Gartner (n.d.) *Gartner glossary: Augmented intelligence*. Gartner. https://www.gartner.com/en/information-technology/glossary/augmented-intelligence

[17] Englebart, D.C. (October 1962). Augmenting human intellect: A conceptual framework. SRI Summary Report AFOSR-3223. https://www.dougengelbart.org/pubs/augment-3906.html

models, algorithms, and data—will become confusing. Because the terms are closely related, we've chosen to focus on the concept of AI models. We want to bring to the fore the idea that every AI model is incomplete, and it's important to know how well the AI model fits the reality we care about, where the model will break down, and how.

Sometimes people avoid talking about the specifics of models to create a mystique. Talking as though AI is unbounded in its potential capabilities and a nearly perfect approximation to reality can convey an excitement about the possibilities of the future. The future, however, can be oversold. Similarly, sometimes people stop calling a model AI when its use becomes commonplace, yet such systems are still AI models with all of the risks discussed here. We need to know exactly when and where AI models fail to align to visions for teaching and learning.

Insight: AI Systems Enable New Forms of Interaction

AI models allow computational processes to make recommendations or plans and also enable them to support forms of interaction that are more natural, such as speaking to an assistant. AI-enabled educational systems will be desirable in part due to their ability to support more natural interactions during teaching and learning. In classic edtech platforms, the ways in which teachers and students interact with edtech are limited. Teachers and students may choose items from a menu or in a multiple-choice question. They may type short answers. They may drag objects on the screen or use touch gestures. The computer provides outputs to students and teachers through text, graphics, and multimedia. Although these forms of inputs and outputs are versatile, no one would mistake this style of interaction with the way two people interact with one another; it is specific to human-computer interaction. With AI, interactions with computers are likely to become more like human-to-human interactions (see Figure 4). A teacher may speak to an AI assistant, and it may speak back. A student may make a drawing, and the computer may highlight a portion of the drawing. A teacher or student may start to write something, and the computer may finish their sentence—as when today's email programs can complete thoughts faster than we can type them.

Additionally, the possibilities for automated actions that can be executed by AI tools are expanding. Current personalization tools may automatically adjust the sequence, pace, hints, or trajectory through learning experiences.[18] Actions in the future might look like an AI system or tool that helps a student with homework[19] or a teaching assistant that reduces a teacher's workload by recommending lesson plans that fit a teacher's needs and are similar to lesson plans a teacher previously liked.[20] Further, an AI-enabled assistant may appear as an additional "partner" in a small group of students who are working together on a collaborative assignment.[21] An AI-enabled tool may also help teachers with complex classroom routines.[22] For example, a

---

[18] Shemshack, A., Spector, J.M. (2020) A systematic literature review of personalized learning terms. *Smart Learning Environments, 7(33).* https://doi.org/10.1186/s40561-020-00140-9

[19] Roschelle, J., Feng, M., Murphy, R. & Mason, C.A. (2016). Online mathematics homework increases student achievement. *AERA Open, 2(4),* 1-12. DOI: 10.1177/2332858416673968

[20] Celik, I., Dindar, M., Muukkonen, H. & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends, 66,* 616–630. https://doi.org/10.1007/s11528-022-00715-y

[21] Chen, C., Park, H.W. & Breazeal, C. (2020). Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Computers & Education, 150,* https://doi.org/10.1016/j.compedu.2020.103836

[22] Holstein, K., McLaren, B.M., & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher–AI complementarity. *Journal of Learning Analytics, 6(2).* https://doi.org/10.18608/jla.2019.62.3

tool may help teachers with orchestrating[23] the movement of students from a full class discussion into small groups and making sure each group has the materials needed to start their work.

Figure 4. Differences that teachers and students may experience in future technologies.

	Familiar Technology Capabilities	Future Technology Capabilities
Input	• Typing	• Speaking
	• Clicking and dragging	• Drawing
	• Touching and gesturing	• Analyzing images and video
Processing	• Displaying information and tasks	• Assisting students and teachers
	• Sequencing learning activities	• Planning and adapting activities
	• Checking student work	• Revealing patterns in student work
Output	• Text	• Conversations
	• Graphics	• Annotating and highlighting
	• Multimedia	• Suggesting and recommending
	• Dashboards	• Organizing and guiding

Key Recommendation: Human in the Loop AI

Many have experienced a moment where technology surprised them with an uncanny ability to recommend what feels like a precisely personalized product, song, or even phrase to complete a sentence in a word processor such as the one being used to draft this document. Throughout this supplement, we talk about specific, focused applications where AI systems may bring value (or risks) into education. At no point do we intend to imply that AI can replace a teacher, a guardian, or an educational leader as the custodian of their students' learning. We talk about the limitations of models in AI and the conversations that educational constituents need to have about what qualities they want AI models to have and how they should be used.

> *"We can use AI to study the diversity, the multiplicity of effective learning approaches and think about the various models to help us get a broader understanding of what effective, meaningful engagement might look like across a variety of different contexts."*
> *—Dr. Marcelo Aaron Bonilla Worsley*

[23] Roschelle, J., Dimitriadis, Y. & Hoppe, U. (2013). Classroom orchestration: Synthesis. *Computers & Education, 69,* 512-526. https://doi.org/10.1016/j.compedu.2013.04.010

These limitations lead to our first recommendation: that we pursue a vision of AI where humans are in the loop. That means that people are part of the process of noticing patterns in an educational system and assigning meaning to those patterns. It also means that teachers remain at the helm of major instructional decisions. It means that formative assessments involve teacher input and decision making, too. One loop is the cycle of recognizing patterns in what students do and selecting next steps or resources that could support their learning. Other loops involve teachers planning and reflecting on lessons. Response to Intervention is another well-known type of loop.

The idea of humans in the loop is part of our broader discussions happening about AI and society, not just AI in education. Interested readers could look for more on human-centered AI, responsible AI, value-sensitive AI, AI for social good, and other similar terms that ally with humans in the loop, such as "human-centered AI."

Exercising judgement and control in the use of AI systems and tools is an essential part of providing the best opportunity to learn for all students—especially when educational decisions carry consequence. AI does not have the broad qualities of contextual judgment that people do. Therefore, people must remain responsible for the health and safety of our children, for all students' educational success and preparation for their futures, and for creating a more equitable and just society.

Learning

The Department's long-standing edtech vision sees students as active learners; students participate in discussions that advance their understanding, use visualizations and simulations to explain concepts as they relate to the real world, and leverage helpful scaffolding and timely feedback as they learn. Constituents want technology to align to and build on these and other research-based understandings of how people learn. Educators can draw upon two books titled *How People Learn* and *How People Learn II* by the National Academies of Sciences, Engineering, and Medicine for a broad synthesis of what we know about learning.[24] As we shape AI-enhanced edtech around research-based principles, a key goal must be to strengthen and support learning for those who have experienced unfavorable circumstances for learning, such as caused by the COVID-19 pandemic or by broader inequities. And we must keep a firm eye toward the forms of learning that will most benefit learners in their future lives in communities and workplaces.

Examples of AI supporting learning principles in this section include the following: AI-based tutoring for students as they solve math problems (based on cognitive learning theories), adapting to learners with special needs (based on the Universal Design for Learning framework and related theories), and AI support for effective student teamwork (based on theories in the field called "Computer Supported Collaborative Learning").

Insight: AI Enables Adaptivity in Learning

Adaptivity has been recognized as a key way in which technology can improve learning.[25] AI can be a toolset for improving the adaptivity of edtech. AI may improve a technology's ability to meet students where they are, build on their strengths, and grow their knowledge and skills. Because of AI's powers of work with natural forms of input and the foundational strengths of AI models (as discussed in the *What is AI?* section), AI can be an especially strong toolkit for expanding the adaptivity provided to students.

And yet, especially with AI, adaptivity is always more specific and limited than what a broad phrase like "meet students where they are" might suggest. Core limits arise from the nature of the model at the heart of any specific AI-enabled system. Models are approximations of reality. When important parts of human learning are left out of the model or less fully developed, the resulting adaptivity will also be limited, and the resulting supports for learning may be brittle or narrow. Consequently, this section on *Learning* focuses on one key concept: Work toward AI models that fit the fullness of visions for learning—and avoid limiting learning to what AI can currently model well.

AI models are demonstrating greater skills because of advances in what are called "large language models" or sometimes "foundational models." These very general models still have limits. For example, generative AI models discussed in the mainstream news can quickly generate convincing essays about a wide variety of topics while other models can draw credible images based on just a few prompts. Despite the excitement about foundational models, experts in our

[24] National Research Council. 2000. *How people learn: Brain, mind, experience, and school.* The National Academies Press. https://doi.org/10.17226/9853; National Academies of Sciences, Engineering, and Medicine. 2018. *How people learn II: Learners, contexts, and cultures*. The National Academies Press. https://doi.org/10.17226/24783

[25] Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction based on adaptive learning technologies. In Mayer, R.E. & Alexander, P.A., *Handbook of research on learning and instruction*, 522-560. ISBN: 113883176X

listening sessions warned that AI models are narrower than visions for human learning and that designing learning environments with these limits in mind remains very important. The models are also brittle and can't perform well when contexts change. In addition, they don't have the same "common sense" judgment that people have, often responding in ways that are unnatural or incorrect.[26] Given the unexpected ways in which foundational models miss the mark, keeping humans in the loop remains highly important.

Intelligent Tutoring Systems: An Example of AI Models

One long-standing type of AI-enabled technology is an Intelligent Tutoring System (ITS).[27] In an early success, scientists were able to build accurate models of how human experts solve mathematical problems. The resulting model was incorporated into a system that would observe student problem solving as they worked on mathematical problems on a computer. Researchers who studied human tutors found that feedback on specific steps (and not just right or wrong solutions) is a likely key to why tutoring is so effective.[28] For example, when a student diverged from the expert model, the system gave feedback to help the student get back on track.[29] Importantly, this feedback went beyond right or wrong, and instead, the model was able to provide feedback on specific steps of a solution process. A significant advancement of AI, therefore, can be its ability to provide adaptivity at the step-by-step level and its ability to do so at scale with modest cost.

As a research and development (R&D) field emerged to advance ITS, the work has gone beyond mathematics problems to additional important issues beyond step-by-step problem solving. In the early work, some limitations can be observed. The kinds of problems that an ITS could support were logical or mathematical, and they were closed tasks, with clear expectations for what a solution and solution process should look like. Also, the "approximation of reality" in early AI models related to cognition and not to other elements of human learning, for example, social or motivational aspects. Over time, these early limitations have been addressed in two ways: by expanding the AI models and by involving humans in the loop, a perspective that is also important now. Today, for example, if an ITS specializes in feedback as a student practices, a human teacher could still be responsible for motivating student engagement and self-regulation along with other aspects of instruction. In other contemporary examples, the computer ITS might focus on problem solving practice, while teachers work with students in small groups. Further, students can be in the loop with AI, as is the case with "open learner models"—a type of AI-enabled system that provides information to support student self-monitoring and reflection.[30]

---

[26] Dieterle, E., Dede, C. & Walker, M. (2022). The cyclical ethical effects of using artificial intelligence in education. *AI & Society*. https://link.springer.com/article/10.1007/s00146-022-01497-w

[27] Mousavinasab, E., Zarifsanaiey, N., R. Niakan Kalhori, S., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments, 29(1),* 142–163. https://psycnet.apa.org/doi/10.1080/10494820.2018.1558257

[28] Van Lehn, K. (2011) The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46(4)*, 197-221. https://doi.org/10.1080/00461520.2011.611369

[29] Ritter, S., Anderson, J.R., Koedinger, K.R. & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review, 14,* 249–255/ https://doi.org/10.3758/BF03194060

[30] Winne, P.H. (2021). Open learner models working in symbiosis with self-regulating learners: A research agenda. *International Journal of Artificial Intelligence in Education, 31(3),* 446-459. https://doi.org/10.1007/s40593-020-00212-4

Although R&D along the lines of an ITS should not limit the view of what's possible, such an example is useful because so much research and evaluation has been done on the ITS approach. Researchers have looked across all the available high-quality studies in a meta-analysis and concluded that ITS approaches are effective.[31] Right now, many school systems are looking at high-intensity human tutoring to help students with unfinished learning. Human tutoring is very expensive, and it is hard to find enough high-quality human tutors. With regard to large-scale needs, if it is possible for an ITS to supplement what human tutors do, it might be possible to extend beyond the amount of tutoring that people can provide to students.

Important Directions for Expanding AI-Based Adaptivity

Adaptivity is sometimes referred to as "personalization." Although this is a convenient term, many observers have noted how imprecise it is.[32] For some educators, personalization means giving learners "voice and choice," and for others it means that a learning management system recommends an individual "playlist" of activities to each student. Hidden in that imprecision is the reality that many edtech products that personalize do so in limited ways. Adjusting the difficulty and the order of lesson materials are among the two most common ways that edtech products adapt. And yet, any teacher knows there is more to supporting learning than adjusting the difficulty and sequence of materials. For example, a good teacher can find ways to engage a student by connecting to their own past experiences and can shape explanations until they really connect in an "aha!" moment for that student. When we say, "meet the learner where they are," human teachers bring a much more complete picture of each learner than most available edtech. The teacher is also not likely to "over personalize" (by performing like an algorithm that only presents material for which the learner has expressed interest), thereby limiting the student's exposure to new topics. The nature of "teachable moments" that a human teacher can grasp is broader than the teachable moments today's AI models grasp.

In our listening sessions, we heard many ways in which the core models in an AI system must be expanded. We discuss these below.

1. **From deficit-based to asset-oriented.** Listening session attendees noted that the rhetoric around adaptivity has often been deficit-based; technology tries to pinpoint what a student is lacking and then provides instruction to fill that specific gap. Teachers also orient to students' strengths; they find competencies or "assets" a student has and use those to build up the students' knowledge. AI models cannot be fully equitable while failing to recognize or build upon each student's sources of competency. AI models that are more asset-oriented would be an advance.
2. **From individual cognition to including social and other aspects of learning.** The existing adaptivity rhetoric has also tended to focus on individualized learning and mostly on cognitive elements of learning, with motivational and other elements only brought in to support the cognitive learning goals. Attendees observe that their vision for learning is broader than cognition. Social learning is important, for example, especially

[31] Kulik, J.A., & Fletcher, J.D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research, 86(1),* 42–78; Ma, W., Adescope, O.O, Nesbit, J.C. & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology, 106(4),* 901–918. http://dx.doi.org/10.1037/a0037123

[32] Plass, J.L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education, 52(3),* 275–300. https://doi.org/10.1080/15391523.2020.1719943;

for students to learn to reason, explain, and justify. For students who are learning English, customized and adaptive support for improving language skills while learning curricular content is clearly important. Developing self-regulation skills is also important. A modern vision of learning is not individualistic; it recognizes that students learn in groups and communities too.

3. **From neurotypical to neurodiverse learners.** AI models could help in including neurodiverse learners (students who access, process, and interact with the world in less common ways than "neurotypical" students) who could benefit from different learning paths and from forms of display and input that fit their strengths. Constituents want AI models that can support learning for neurodiverse learners and learners with disabilities. Thus, they want AI models that can work with multiple paths to learning and multiple modalities of interaction. Such models should be tested for efficacy, to guard against the possibility that some students could be assigned a "personalized" but inadequate learning resource. In addition, some systems for neurodiverse students are presently underutilized, so designs that support intended use will also be important.

4. **From fixed tasks to active, open, and creative tasks.** As mentioned above, AI models are historically better at closed tasks like solving a math problem or logical tasks like playing a game. In terms of life-wide and lifelong opportunities, we value learning how to succeed at open-ended and creative tasks that require extended engagement from the learner, and these are often not purely mathematical or logical. We want students to learn to invent and create innovative approaches. We want AI models that enable progress on open, creative tasks.

5. **From correct answers to additional goals.** At the heart of many adaptivity approaches now on the market, the model inside the technology counts students' wrong answers and decides whether to speed up, slow down, or offer a different type of learning support. Yet, right and wrong answers are not the only learning goals. We want students to learn how to self-regulate when they experience difficulties in learning, for example, such as being able to persist in working on a difficult problem or knowing how and when to ask for help. We want learners to become skilled in teamwork and in leading teams. As students grow, we want them to develop more agency and to be able to act on their own to advance toward their own learning goals.

Listing every dimension of expansion that we heard in our listening sessions is beyond the scope of this report. Some additional dimensions are presented in the following sections on *Teaching, Assessment,* and *Research*. For example, in *Research*, we discuss all the ways in which AI systems have trouble with context—context that humans readily grasp and consider.

Overall, constituents in the listening sessions realized we need an ambitious outlook on learning to respond to the future today's learners face. Constituents were concerned about ways in which AI might narrow learning. For example, if the incorporation of AI into education slowed attention to students' skills on creative, open-ended tasks and their ability to lead and collaborate in teams, then school districts may be less able to realize their students' progress in relation to a Portrait of a Graduate who excels in communication and other skills valued in communities and careers.

Constituents reminded us that as we conceptualize what we want AI in edtech to accomplish, we must start and constantly revisit a human-centered vision of learning.

A Duality: Learning With and About AI

As AI is brought into schools, two broad perspectives about AI in education arise: (1) AI in support of student learning; and (2) support for learning about AI and related technologies. So far, we've discussed AI systems and tools to support student learning and mastery of subjects like mathematics and writing. Yet, it is also important that students learn about AI, critically examine its presence in education and society, and determine its role and value in their own lives and careers. We discuss risks across each section in this report. Here, it is important for students to become more aware of and savvy to the risks of AI—including risks of bias and surveillance—as they appear in all elements of their lives. In the recent past, schools have supported students' understanding of cybersecurity, for example. AI will bring new risks, and students need to learn about them.

We are encouraged by efforts we've seen underway that would give students opportunities to learn about how AI works while also giving them opportunities to discuss relevant topics like privacy and security.[33] Other learning goals are noted in the K-12 Computer Science Framework. We've seen that students can begin learning about AI in elementary, middle, and high school. They can use AI to design simulations and products that they find exciting. And we've seen that students want to talk about the ethics of products they experience in their everyday lives and have much to say about the kinds of products they'd like to see or not see in school. (And later, in the *Research* section, we note the desire for co-design processes that involve students in creating the next generation of AI-enabled edtech). Overall, it's important to balance attention to using AI to support learning and giving students opportunities to learn about AI.

A Challenge: Systems Thinking About AI in Education

As AI expands into the educational system, our listening session attendees reminded us that it will be entering parts or locations of the system that are presently dysfunctional. AI is certainly not a fix for broken systems, and instead, must be used with even more care when the systems' context is unstable or uncertain.

[33] Forsyth, S., Dalton, B., Foster, E.H., Walsh, B., Smilack, J., & Yeh, T. (2021, May). Imagine a more ethical AI: Using stories to develop teens' awareness and understanding of artificial intelligence and its societal impacts. In 2021 Conference on Research in Equitable and Sustained Participation in Engineering, Computing, and Technology (RESPECT). IEEE. https://doi.org/10.1109/RESPECT51740.2021.9620549; Zhang, H., Lee, I., Ali, S., DiPaola, D., Cheng, Y., & Breazeal, C. (2022). Integrating ethics and career futures with technical learning to promote AI literacy for middle school students: An exploratory study. *International Journal of Artificial Intelligence in Education, 1–35.* https://doi.org/10.1007/s40593-022-00293-3

"First and foremost, they are getting deployed in educational contexts that are already fragmented and broken and unequal. Technology doesn't discriminate—we do. So, as we think about the application of these new systems, we have to really think about the contextual application of AI."
—Dr. Nicole Turner

As discussed previously, because AI systems and tools do not fully align with goals for learning, we have to design educational settings to situate AI in the right place, where educators and other adults can make effective use of these tools for teaching and learning. Within the ITS example, we saw that AI could make learning by practicing math problems more effective, and a whole curricular approach might include roles for teachers that emphasize mathematical practices like argumentation and modeling. Further, small-group work is likely to remain important: Students might work in small groups to use mathematics to predict or justify as they work on responding to a realistic challenge. At the present, one "right place" for people, and not AI, is understanding how learning can be culturally responsive and culturally sustaining, as AI is not even close to being ready to connect learning to the unique strengths in a student's community and family.

Open Questions About AI for Learning

With advances occurring in the foundations for AI, opportunities to use AI in support of learning are rapidly expanding. As we explore these opportunities, the open questions below deserve ongoing attention:

- To what extent is AI enabling adaptation to students' strengths and not just deficits? Is AI enabling improved support for learners with disabilities and English language learners?
- How are youth voices involved in choosing and using AI for learning?
- Is AI leading to narrower student activities (e.g., procedural math problems), or the fuller range of activities highlighted in the National Educational Technology Plan (NETP), which emphasizes features such as personalized learning, project-based learning, learning from visualizations, simulations, and virtual reality, as well as learning across school, community, and familial settings?
- Is AI supporting the whole learner, including social dimensions of learning such as enabling students to be active participants in small group and collaborative learning? For example, does AI contribute to aspects of student collaboration we value like shared attention, mutual engagement, peer help, self-regulation, and building on each other's contributions?
- When AI is used, are students' privacy and data protected? Are students and their guardians informed about what happens with their data?
- How strong are the processes or systems for monitoring student use of AI for barriers, bias, or other undesirable consequences of AI use by learners? How are emergent issues addressed?
- Is high-quality research or evaluations about the impacts of using the AI system for student learning available? Do we know not only whether the system works but for whom and under what conditions?

Key Recommendation: Seek AI Models Aligned to a Vision for Learning

We've called attention to how advances in AI are important to adaptivity but also to ways in which adaptivity is limited by the model's inherent quality. We noted that a prior wave of edtech used the term "personalized" in differing ways, and it was often important to clarify what personalization meant for a particular product or service. Thus, our key recommendation is to tease out the strengths and limitations of AI models inside forthcoming edtech products and to focus on AI models that align closely to desired visions of learning. AI is now advancing rapidly, and we should differentiate between products that have simple AI-like features inside and products that have more sophisticated AI models.

Looking at what's happening in research and development, we can see significant effort and push toward overcoming these limitations. We noted that decision makers need to be careful about selecting AI models that might narrow their vision for learning, as general artificial intelligence does not exist. And because AI models will always be narrower than real world experience, we need to proceed with systems thinking in which humans are in the loop, with the strengths and weaknesses of the specific educational system considered. We hold that the full system for learning is broader than its AI component.

Teaching

Teachers have long envisioned many things that technology could make possible for teachers, their classrooms, and their students but not the changes wrought by the recent pandemic. Today, nearly all teachers have experienced uses of technologies for instruction that no one anticipated. Some of those experiences were positive, and others were not. All of the experiences provide an important context as we think further about teaching and technology.

There is a critical need to focus on addressing the challenges teachers experience. It must become easier for teachers to do the amazing work they always do. We must also remember why people choose the teaching profession and ensure they can do the work that matters. This section discusses examples of AI supporting teachers and teaching including these concepts: AI assistants to reduce routine teaching burdens; AI that provides teachers with recommendations for their students' needs and extends their work with students; and AI that helps teachers to reflect, plan, and improve their practice.

> *"One opportunity I see with AI is being able to reduce the amount of attention I have to give to administrative things and increase the amount of attention I can give to my students with their learning needs in the classroom. So that's the first one that I'd say that I'm super excited about the possibility of AI to support me as a teacher."*
> *—Vidula Plante*

Always Center Educators in Instructional Loops

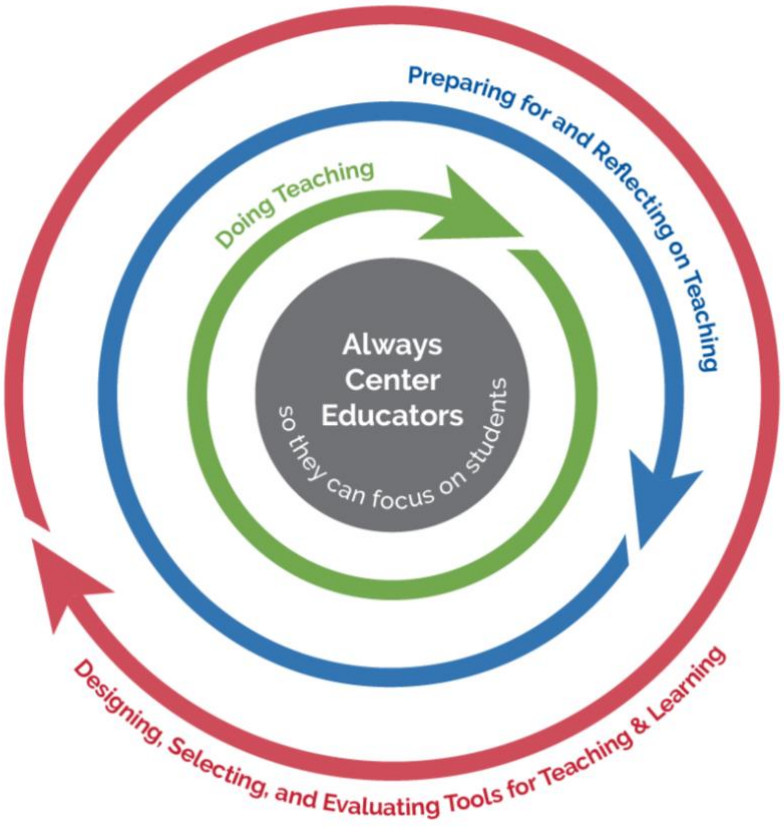
To succeed with AI as an enhancement to learning and teaching, we need to always center educators (ACE). Practically speaking, practicing "ACE in AI" means keeping a humanistic view of teaching front and center. ACE leads the Department to confidently respond "no" when asked "will AI replace teachers?" ACE is not just about making teachers' jobs easier but also making it possible to do what most teachers want to do. That includes, for example, understanding their students more deeply and having more time to respond in creative ways to teachable moments.

To bring more precision to how and where we should center educators, we return to our advocacy for human in the loop AI and ask, what are the loops in which teachers should be centered? Figure 5 suggests three key loops (inspired by research on adaptivity loops[34]):

[34] Aleven, V., McLaughlin, E.A., Glenn, R.A., & Koedinger, K.R. (2016). Instruction based on adaptive learning technologies. In Mayer, R.E. & Alexander, P.A., *Handbook of research on learning and instruction*, 522-560. ISBN: 113883176X

1. The loop in which teachers make moment-to-moment decisions as they do the immediate work of teaching.
2. The loop in which teachers prepare for, plan, and reflect on teaching, which includes professional development.
3. The loop in which teachers participate in decisions about the design of AI-enabled technologies, participate in selecting the technologies, and shape the evaluation of technologies—thus setting a context for not only their own classroom but those of fellow teachers as well.

Figure 5: Three ways to center educators as we conceptualize human in the loop AI



Please note that in the next section, on *Formative Assessment*, we also discuss teachers' important role in feedback loops that support students and enable school improvement. That section also includes a discussion of the concepts of "bias" and "fairness," which are important to teachers.

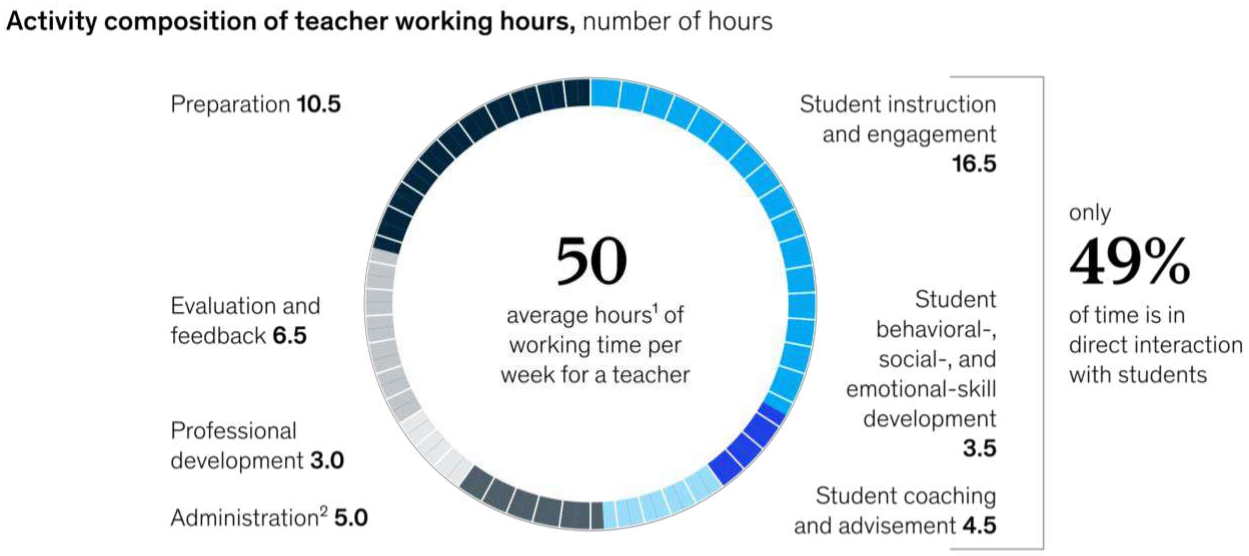
Insight: Using AI to Improve Teaching Jobs

The job of teaching is notoriously complex, with teachers making thousands of decisions each day. Teachers participate in classroom processes, in interactions with students beyond classrooms, in work with fellow teachers, and in administrative functions. They also are part of their communities and thus are expected to interact with families and caregivers.

If the teacher is able to efficiently predict and understand the range of other answers given by students in the class, it becomes possible to think creatively about the novel answer and figure how and why the student might have generated it.[35]

We think about how much easier some everyday tasks have become. We can request and receive alerts and notifications about events. Selecting music that we want to hear used to be a multistep process (even with digital music), and now we can speak the name of a song we want to hear, and it plays. Likewise, mapping a journey used to require a cumbersome study of maps, but now cell phones let us choose among several transportation options to reach a destination. Why can't teachers be supported to notice changing student needs and provided with supports to enact a technology-rich lesson plan? Why can't they more easily plan their students' learning journeys? When things change in a classroom, as they always do, why don't the tools of the classroom make it easier for teachers to adapt to student strengths and needs on the fly?

Figure 6: Teachers work about 50 hours a week, spending less than half the time in direct interaction with students.



A report by McKinsey[36] first suggested that AI's initial benefit could be to improve teaching jobs by reducing low-level burdens in administrative or clerical work (Figure 6). The report also suggests that recovered time from AI-enabled technology should be rededicated toward more

[35] Hammerness, K., Darling-Hammond, L., & Bransford, J. (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. Jossey-Bass. ISBN: 0787996343

[36] Bryant, J., Heitz,C., Sanghvi, S., & Wagle, D. (2020, January 14). *How artificial intelligence will impact K-12 teachers.* McKinsey. https://www.mckinsey.com/industries/education/our-insights/how-artificial-intelligence-will-impact-k-12-teachers

effective instruction—particularly, outcomes such as reducing the average 11 hours of weekly preparation down to only six. We highlight these opportunities and two others below.

1. **Handling low-level details to ease teaching burdens and increase focus on students.** A good teacher must master all levels of details, big and small. When working with a particular student, the teacher may wish to later send that student a helpful learning resource. How will they remember to send it? A voice assistant or other forms of an AI assistant could make it easier to stay organized by categorizing simple voice notes for teachers to follow up on after a classroom session ends. We are beginning to see AI-enabled voice assistants in the market, and they could do many simple tasks so that the teachers can stay focused on students. These tasks can include record-keeping, starting and stopping activities, controlling displays, speakers, and other technologies in the classroom, and providing reminders. Many workers may eventually use assistants to make their jobs easier, and teachers are the most deserving of efforts to ease their jobs now.

2. **Extending beyond the teacher's availability with their students but continuing to deliver on the teacher's intent.** Teachers almost always want to do more with each student than they can, given the limited number of hours before the next school day. A teacher may wish to sit with the student as they practice 10 more math problems, giving them ongoing support and feedback. If the teacher can sit with the student for only three problems, perhaps they could delegate to an AI-enabled learning system to help with the rest. Teachers cannot be at their best if on call at all hours to help with homework, but perhaps they can indicate what types of supports, hints, and feedback they want students to receive while studying after school hours. An AI assistant can ensure that students have that support wherever and whenever they do homework or practice skills on their own. Teachers may wish to provide more extensive personal notes to families/caregivers, and perhaps an AI assistant could help with drafts based on students' recent classroom work. Then, the teacher could review the AI-generated comments and quickly edit where needed before returning it to the student for another draft. AI tools might also help teachers with language translation so they can work with all parents and caregivers of their students. AI tools might also help teachers with awareness. For example, in the next section, *Formative Assessment*, we note that teachers can't always know what's going on for each student and in each small group of students; emerging products might signal to the teacher when a student or teacher may need some more personal attention.

3. **Making teacher professional development more productive and fruitful.** Emerging products already enable a teacher to record her classroom and allow an AI algorithm to suggest highlights of the classroom discussion worth reviewing with a professional development coach.[37] AI can compute metrics, such as whether students have been talking more or less, which are difficult for a teacher to calculate during a lesson.[38] For

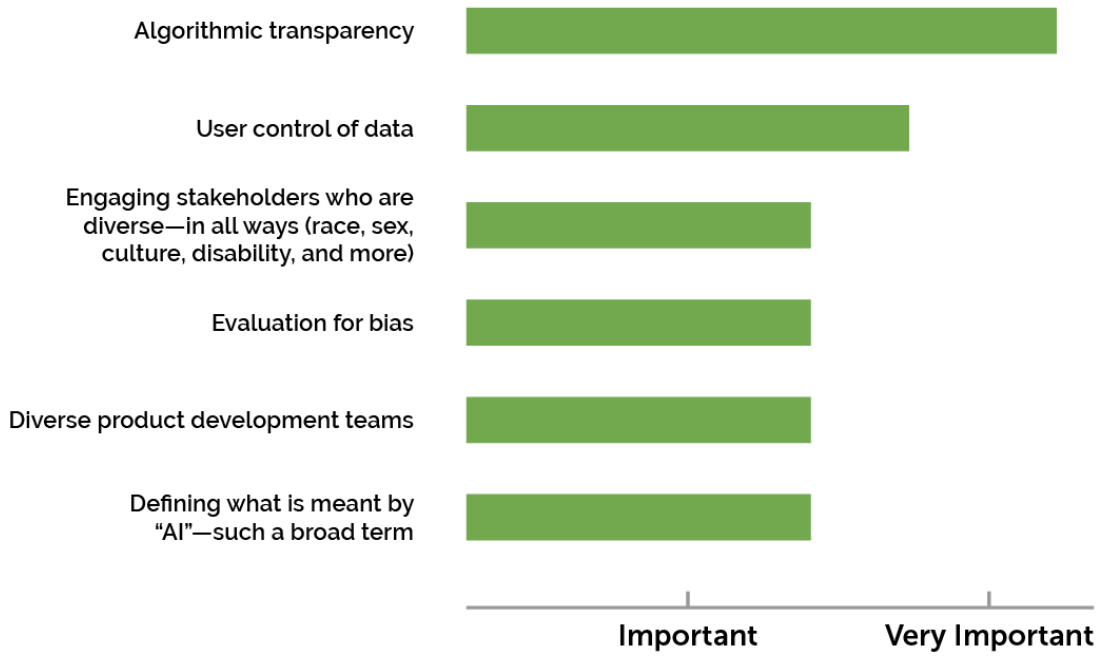
[37] Chen, G., Clarke, S., & Resnick, L.B. (2015). Classroom Discourse Analyzer (CDA): A discourse analytic tool for teachers. *Technology, Instruction, Cognition and Learning, 10(2),* 85-105

[38] Jensen, E., Dale, M., Donnelly, P.J., Stone, C., Kelly, S., Godley, A. & D'Mello, S.K. (2020). Toward automated feedback on teacher discourse to enhance teacher learning. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). https://doi.org/10.1145/3313831.3376418

teachers who want to increase student engagement, these metrics can be a valuable tool. Classroom simulation tools are also emerging and can enable teachers to practice their skills in realistic situations.[39] Simulators can include examples of teaching from a real classroom while changing the faces and voices of the participants so that teaching situations can be shared and discussed among teachers without revealing identities.

Note the emphasis above on what listening-session panelist Sarah Hampton said about the human touch. Teachers will feel that AI is helping them teach with a focus on their human connection to their students when the necessary (but less meaningful) burdens of teaching are lessened. In Figure 7, below, see concerns that teachers raised about AI during listening sessions.

Figure 7: Concerns raised during the listening session about teaching with AI



Preparing and Supporting Teachers in Planning and Reflecting

ACE also means preparing teachers to take advantage of possibilities like those listed above and more. In the *Research* section, we highlight how pre-service education still tends to compartmentalize and inadequately address the topic of technology. That section suggests a need to invest in research about how to deeply integrate technology in pre-service teacher training programs. In-service teachers, too, will need professional development to take advantage of opportunities that AI can provide, like those presented in the *Teaching* section. Professional development will need to be balanced not only to discuss opportunities but also to inform teachers of new risks, while providing them with tools to avoid the pitfalls of AI.

[39] Ersozlu, Z., Ledger, S., Ersozlu, A., Mayne, F., & Wildy, H. (2021). Mixed-reality learning environments in teacher education: An analysis of TeachLivETM Research. *SAGE Open, 11(3).* https://doi.org/10.1177/21582440211032155.

"Humans are well suited to discern the outcomes…because we are the ones that have the capacity for moral reflection and empathy. So, in other words, I want the AI to help me really quickly and easily see what my student needs in their learning journey."
—Sarah Hampton

By nature, teaching requires significant time in planning as well to account for the breadth of needs across their rosters—especially for inclusive learning environments and students with IEPs and 504 plans. AI could help teachers with recommendations that are tuned to their situation and their ways of practicing teaching and support with adapting found materials to fit their exact classroom needs. For students with an IEP, AI could help with finding components to add to lesson plans to fully address standards and expectations and to meet each student's unique requirements. Even beyond finding components, AI might help adapt standardized resources to better fit specific needs—for example, providing a voice assistant that allows a student with a visual difficulty to hear material and respond to it or permitting a group of students to present their project using American Sign Language (ASL) which could be audibly voiced for other students using an AI ASL-to-Spoken-English translation capability. Indeed, coordinating IEPs is time-consuming work that might benefit from supportive automation and customized interactivity that can be provided by AI.

Reflection is important too. In the bustle of a classroom, it is sometimes difficult to fully understand what a student is expressing or what situations lead to certain positive or negative behaviors. Again, context is paramount. In the moment, teachers may not be aware of external events that could shape their understanding of how students are showing up in their classrooms. Tools that notice patterns and suggest ways to share information might help students and teachers communicate more fully about strengths and needs.

Designing, Selecting, and Evaluating AI Tools

The broadest loop teachers should be part of is the loop that determines what classroom tools do and which tools are available. Today, teachers already play a role in designing and selecting technologies. Teachers can weigh in on usability and feasibility. Teachers examine evidence of efficacy and share their findings with other school leaders. Teachers already share insights on what is needed to implement technology well.

While these concerns will continue, AI will raise new concerns too. For example, the following *Formative Assessment* section raises concerns about bias and fairness that can lead to algorithmic discrimination. Those concerns go beyond data privacy and security; they raise attention to how technologies may unfairly direct or limit some students' opportunities to learn. A key takeaway here is that teachers will need time and support so they can stay abreast of both the well-known and the newer issues that are arising and so they can fully participate in design, selection, and evaluation processes that mitigate risks.

Challenge: Balancing Human and Computer Decision-Making

One major new challenge with AI-enabled tools for teachers is that AI can enable autonomous activity by a computer, and thus when a teacher delegates work to an AI-enabled tool, it may

carry on with that work somewhat independently. Professor Inge Molenaar[40] has wondered about the challenges of control in a hybrid teaching scenario: When should a teacher be in control? What can be delegated to a computational system? How can a teacher monitor the AI system and override its decisions or take back control as necessary?

Figure 8: The tension between human and AI decision making: Who is in control?

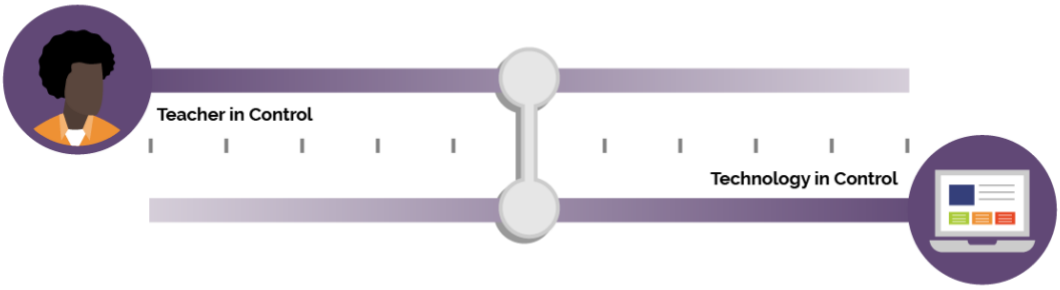


Figure 8 expresses the tension around control. To the left, the teacher is fully in control, and there is no use of AI in the classroom. To the right, the technology is fully in control with no teacher involved—a scenario which is rarely desirable. The middle ground is not one dimensional and involves many choices. Molenaar analyzed products and suggests some possibilities:

- The technology only offers information and recommendations to the teacher.
- The teacher delegates specific types of tasks to the technology, for example, giving feedback on a particular math assignment or sending out reminders to students before an assignment is due.
- The teacher delegates more broadly to the technology, with clear protocols for alerts, for monitoring, and for when the teacher takes back control.

These and other choices need to be debated openly. For example, we may want to define instructional decisions that have different kinds of consequences for a student and be very careful about delegating control over highly consequential decisions (for example, placement in a next course of study or disciplinary referrals). For human in the loop to become more fully realized, AI technologies must allow teacher monitoring, have protocols to signal a teacher when their judgment is needed, and allow for classroom, school, or district overrides when they disagree with an instructional choice for their students. We cannot forget that if a technology allows a teacher choice—which it should—it will take significant time for a teacher to think through and set up all the options, requiring greater time initially.

Challenge: Making Teaching Jobs Easier While Avoiding Surveillance

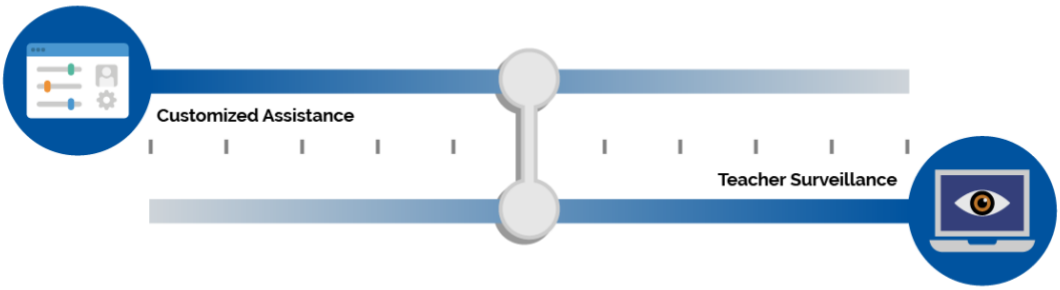
We also recognize that the very technologies that make jobs easier might also introduce new possibilities for surveillance (Figure 9). In a familiar example, when we enable a voice assistant in the kitchen, it might help us with simple household tasks like setting a cooking timer. And yet the same voice assistant might hear things that we intended to be private. This kind of dilemma will

[40] Molenaar, I. (2022). Towards hybrid human-AI learning technologies. European Journal of Education, 00, 1–14. https://doi.org/10.1111/ejed.12527

occur in classrooms and for teachers. When they enable an AI-assistant to capture data about what they say, what teaching resources they search for, or other behaviors, the data could be used to personalize resources and recommendations for the teacher. Yet the same data might also be used to monitor the teacher, and that monitoring might have consequences for the teacher. Achieving trustworthy AI that makes teachers' jobs better will be nearly impossible if teachers experience increased surveillance.

A related tension is that asking teachers to be "in the loop" could create more work for teachers if not done well, and thus, being in the loop might be in tension with making teaching jobs easier. Also related is the tension between not trusting AI enough (to obtain assistance) or trusting it too much (and incurring surveillance or loss of privacy). For example, researchers have documented that people will follow instructions from a robot during a simulated fire emergency even when (a) they are told the robot is broken and (b) the advice is obviously wrong.[41] We anticipate teachers will need training and support to understand how and when they will need to exercise human judgement.

Figure 9: Highly customized assistance vs. increased teacher surveillance



Challenge: Responding to Students' Strengths While Protecting Their Privacy

Educators seek to tackle inequities in learning, no matter how they manifest locally (e.g. in access to educational opportunities, resources, or supports). In culturally responsive[42] and culturally sustaining[43] approaches, educators design materials to build on the "assets"—individual, community, and cultural strengths that students bring to learning. Along with considering assets, of course, educators must meet students where they are, including both strengths and needs. AI could assist in this process by helping teachers with customizing curricular resources, for example. But to do so, the data inputted in an AI-enabled system would have to provide more information about the students. This information could be, but need not be, demographic details. It could also be information about students' preferences, outside interests, relationships,

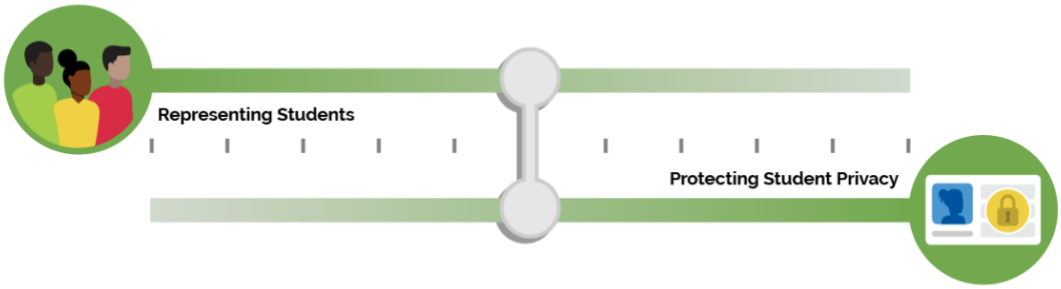
[41] Wagner, A.R., Borenstein, J. & Howard, A. (September 2018). Overtrust in the robotics age. *Communications of the ACM, 61(9),*22-24. https://doi.org/10.1145/3241365

[42] Gay, G. (2018). *Culturally responsive teaching: Theory, research, and practice.* Teachers College Press. ISBN: 978-0807758762

[43] Paris, D., & Alim, H.S. (Eds.). (2017). Culturally sustaining pedagogies: Teaching and learning for justice in a changing world. Teachers College Press. ISBN: 978-0807758342

or experiences.[44] What happens to this data, how it is deleted, and who sees it is of huge concern to educators. As educators contemplate using AI-enabled technologies to assist in tackling educational inequities, they must consider whether the information about students shared with or stored in an AI-enabled system is subject to federal or state privacy laws, such as FERPA. Further, educators must consider whether interactions between students and AI systems create records that must be protected by law, such as when a chatbot or automated tutor generates conversational or written guidance to a student. Decisions made by AI technologies, along with explanations of those decisions that are generated by algorithms may also be records that must be protected by law. Therein, a third tension emerges, between more fully representing students and protecting their privacy (Figure 10).

Figure 10: Responding to students' strengths while fully protecting student privacy



Further, representation would be just a start toward a solution. As discussed earlier in this report, AI can introduce algorithmic discrimination through bias in the data, code, or models within AI-enhanced edtech. Engineers develop the pattern detection in AI models using existing data, and the data they use may not be representative or may contain associations that run counter to policy goals. Further, engineers shape the automations that AI implements when it recognizes patterns, and the automations may not meet the needs of each student group with a diverse population. The developers of AI are typically less diverse than the populations they serve, and as a consequence, they may not anticipate the ways in which pattern detection and automation may harm a community, group, or individual.

AI could help teachers to customize and personalize materials for their students, leveraging the teacher's understanding of student needs and strengths. It is time consuming to customize curricular resources, and teachers are already exploring how AI chatbots can help them design additional resources for their students. An elementary school teacher could gain powerful supports for changing the visuals in a storybook to engage their students or for adapting language that poorly fits local manners of speaking or even for modifying plots to incorporate other dimensions of a teacher's lesson. In the *Learning* section, we noted that AI could help identify learner strengths. For example, a mathematics teacher may not be aware of ways in which a student is making great sense of graphs and tables about motions when they are in another teacher's physics classroom and might not realize that using similar graphs about

[44] Zacamy, J. & Roschelle, J. (2022). Navigating the tensions: How could equity-relevant research also be agile, open, and scalable? Digital Promise. http://hdl.handle.net/20.500.12265/159; Baker, R.S., Esbenshade, L., Vitale, J., & Karumbaiah, S. (2022). Using demographic data as predictor variables: A questionable choice. https://doi.org/10.35542/osf.io/y4wvj

motion could help with their linear function lesson. AI might help teachers when they seek to reflect student strengths by creating or adapting instructional resources.

Yet, the broad equity challenges of avoiding algorithmic discrimination while increasing community and cultural responsiveness must be approached within the four foundations we earlier outlined: human in the loop, equity, safety and effectiveness, and evaluation of AI models. We cannot expect AI models to respect cultural responsiveness. The Department is particularly concerned that equity is something that engaged educators and other responsive adults are in the best position to address and something that is never solely addressable as a computational problem.

Questions Worth Asking About AI for Teaching

As leaders in both pre-service and post-service teacher education contemplate how AI can improve teaching (along with policymakers, developers, and researchers), we urge all in the ecosystem to spend more time asking these questions:

- Is AI improving the quality of an educator's day-to-day work? Are teachers experiencing less burden and more ability to focus and effectively teach their students?
- As AI reduces one type of teaching burden, are we preventing new responsibilities or additional workloads being shifted and assigned to teachers in a manner that negates the potential benefits of AI?
- Is classroom AI use providing teachers with more detailed insights into their students and their strengths while protecting their privacy?
- Do teachers have oversight of AI systems used with their learners? Are they exercising control in the use of AI-enabled tools and systems appropriately or inappropriately yielding decision-making to these systems and tools?
- When AI systems are being used to support teachers or to enhance instruction, are the protections against surveillance adequate?
- To what extent are teachers able to exercise voice and decision-making to improve equity, reduce bias, and increase cultural responsiveness in the use of AI-enabled tools and systems?

Key Recommendation: Inspectable, Explainable, Overridable AI

In the Introduction, we discuss the notion that when AI is incorporated into a system, the core of the AI is a model. In the *Learning* section, we discuss that we need to be careful that models align to the learning we envision (e.g., that they aren't too narrow). Now, based on the needs of teachers (as well as students and their families/caregivers), we add another layer to our criteria for good AI models: the need for explainability.[45] Some AI models can recognize patterns in the world and do the right action, but they cannot explain why (e.g., how they arrived at the

[45] Khosravi, H., Shum, S.B., Chen, G, Conati, C., Tsai,Y-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence, 3.* https://doi.org/10.1016/j.caeai.2022.100074

connection between the pattern and the action). This lack of explainability will not suffice for teaching; teachers will need to know how an AI model analyzed the work of one of their students and why the AI model recommended a particular tutorial, resource, or next step to the student.

Thus, explainability of an AI system's decision is key to a teacher's ability to judge that automated decision. Such explainability helps teachers to develop appropriate levels of trust and distrust in AI, particularly to know where the AI model tends to make poor decisions. Explainability is also key to a teacher's ability to monitor when an AI system may be unfairly acting on the wrong information (and thus may be biased. We discuss bias and fairness more in the *Assessment* section next).

Surrounding the idea of explainability is the need for teachers to be able to inspect what an AI model is doing. For example, what kinds of instructional recommendations are being made and to which students? Which students are being assigned remedial work in a never ended loop? Which are making progress? Dashboards in current products present some of this information, but with AI, teachers may want to further explore which decisions are being made and for whom and know of the student-specific factors that an AI model had available (and possibly which factors were influential) when reaching a particular decision. For example, some of today's adaptive classroom products use limited recommendation models that only consider student success on the last three mathematics problems and do not consider other variables that a teacher would know to consider, such as whether a student has an IEP Plan or other needs.

Our call for attending to equity considerations as we evaluate AI models requires information about how discriminatory bias may arise in particular AI systems and what developers have done to address it. This can only be achieved with transparency for how the tools use datasets to achieve outcomes and what data they have available or that a teacher could include in her judgement but are not available to the system (IEP status is offered as an example above).

Teachers will also need the ability to view and make their own judgement about automated decisions, such as decisions about which set of mathematics problems a student should work on next. They need to be able to intervene and override decisions when they disagree with the logic behind an instructional recommendation.[46] Teachers need protection against adverse ramifications when they assert human judgement over an AI system's decision.

[46] Ruiz, P. & Fusco, J. (2022). *Teachers partnering with artificial intelligence: Augmentation and automation.* Digital Promise. https://digitalpromise.org/2022/07/06/teachers-partnering-with-artificial-intelligence-augmentation-and-automation/

"These systems sometimes are seen as a black box kind of a situation where predictions are made based on lots of data. But what we need is to have a clear view—to clearly show how those recommendations or those interactions are made and what evidence is used or what data is used to be able to make those recommendations so teachers and everyone involved know about why that kind of system is providing that type of information. So, having open learning environments or inspectable learner models or applications where the stakeholders can understand how these systems make decisions or recommendations is going to be an important aspect in the future of teaching and learning."
—Diego Zapata-Rivera

Formative Assessment

Formative assessment is traditionally a key use of edtech because feedback loops are vital to improving teaching and learning.[47] As we have emphasized throughout this report, a top priority with AI is to keep humans in the loop and in control, which includes focusing on the people engaged with formative assessments: students, teachers, school leaders, families/caregivers, and others who support learners. In the definition below, please note the overlap between definitions of AI and formative assessment; both have to do with detecting patterns and choosing a future course of action (that adapts to learner strengths and needs).

> *Assessment refers to all those activities undertaken by teachers, and by the students in assessing themselves, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged. Such assessment becomes "formative assessment" when the evidence is actually used to adapt the teaching to meet the needs.*[48]

Building on Best Practices

A number of dimensions hold potential for shaping the future of formative assessments, and many have ready extensions to the field of AI-enabled systems and tools. For example, the 2017 NETP discussed how technology can lead to improved formative assessments along seven dimensions, listed below:

1. **Enabling Enhanced Question Types:**
   to give students more ways to show what they know and can do.
2. **Measurement of Complex Competencies:**
   to better elicit growth in important skills that go beyond typical subject matter standards, for example, in measuring practices, social skills like teamwork, self-regulation, and work-relevant skills (e.g., making presentations or leading teams).
3. **Providing Real-Time Feedback:**
   to maintain and increase student engagement and to support effective learning, providing timely and helpful responses and suggestions to each learner.
4. **Increasing Accessibility:**
   to include neurodiverse learners and to engage learners' best communication capabilities as they share what they know and can do.

---

[47] Shute, V.J. (2008). Focus on formative feedback. *Review of Educational Research, 78(1),* 153–189. https://doi.org/10.3102/0034654307313795

[48] Black, P. & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. Phi Delta Kappan, 92(1), 81-90. https://kappanonline.org/inside-the-black-box-raising-standards-through-classroom-assessment/

5. **Adapting to Learner Ability and Knowledge:**
   to make assessments more precise and efficient.
6. **Embedded Assessment in the Learning Process:**
   to emphasize an assessment's role in improving teaching and learning (this report does not focus on assessment for accountability purposes).
7. **Assess for Ongoing Learning:**
   to reveal progress over time and not just predetermined milestones.

AI models and AI-enabled systems may have potential to strengthen formative assessments. In one example, a question type that invites students to draw a graph or create a model can be analyzed with AI algorithms,[49] and similar student models might be grouped for the teacher to interpret. Enhanced formative assessment may enable teachers to better respond to students' understanding of a concept like "rate of change" in a complex, real-world situation. AI can also give learners feedback on complex skills, such as learning American Sign Language[50] or speaking a foreign language[51] and in other practice situations where no person is available to provide immediate feedback.

Generally, an AI assistant may be able to reduce the load for teachers related to grading simpler aspects of student responses, allowing the teacher to focus their specialized judgment on important qualities of a whole essay or a complex project. We also may be able to better provide feedback with accessibility. For example, an AI-enabled learning technology may be able to interact verbally with a student about their response to an essay prompt, asking questions that guide the student to clarify their argument without requiring the student to read a screen or type at a keyboard. In the examples shared earlier in the *Learning* section, we also see that AI can be embedded in the learning process, providing feedback to students as they work to solve a problem, rather than only later after the student has reached a wrong answer. When formative assessment is more embedded, it can better support learning, and timely feedback is critical.[52]

Although there are many points of connection like these between AI and formative assessments, our listening sessions also revealed attendees' desire to tackle some existing shortcomings in the field of formative assessment; namely, the time-consuming and sometime onerous nature of taking tests, quizzes, or other assessments and the lack of perceived value in the feedback loop by teachers and students.

Implications for Teaching and Learning

Real-time instructional feedback can be beneficial when it helps learners and teachers to improve. But common experience too often leaves students and teachers with unpleasant feelings toward assessment and thus poses a provocative conflict between the potential benefits

---

[49] Zhai, X., He, P., Krajcik, J. (2022). Applying machine learning to automatically assess scientific models. *Journal of Research in Science Teaching*. https://doi.org/10.1002/tea.21773

[50] Shao, Q., Sniffen, A., Blanchet, J., Hillis, M.E., Shi, X., Haris, T.K., & Balkcom, D. (2020). Teaching american sign language in mixed reality. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(4),* 1-27. https://doi.org/10.1145/3432211

[51] Godwin-Jones, R. (2021). Big data and language learning: Opportunities and challenges. *Language Learning & Technology, 25(1),* 4–19. http://hdl.handle.net/10125/44747

[52] Wiggins, G. (2015). *Seven keys to effective feedback*. ACSD. https://www.ascd.org/el/articles/seven-keys-to-effective-feedback

of data collected through formative assessments and the practical implications of administering additional assessments in classrooms and schools.

Some AI-enabled systems and tools seek to address this potential conflict. For example, one AI-enabled reading tutor listens to students as they read aloud and provides on-the-spot feedback to improve their reading.[53] Students reportedly enjoyed reading aloud, and the approach was effective. Researchers have also embedded formative assessments in games so that students can show how well they understand Newtonian physics as they play increasingly difficult levels of a game.[54] If a student can more easily ask for and receive help when they feel frustrated or confused, reducing those feelings can feel encouraging. Student feelings of safety, confidence, and trust in the feedback generated by these AI-enabled systems and tools are essential to showcase their learning. That focus on learning growth and gains is optimal (absent negative consequences or a high-stakes environment).[55]

AI-enhanced formative assessments may have the potential to save teachers' time (e.g., time spent on grading), allowing the instructor to spend more time engaged in helping students. AI-enhanced assessments may also benefit teachers if they provide detailed insights about student strengths or needs that may not be visible and if they support instructional adaptation or improvement by suggesting a small set of evidence-based recommendations for helping students master content. Such assessments may also be helpful outside of the classroom if it can provide feedback when the teacher is not available, for example, in completing homework or practicing a concept during study hall. As we discussed in the *Teaching* section, an essential aspect of deploying AI-based formative assessment must be centering teachers in system design.

Insight: AI Can Enhance Feedback Loops

The term "formative assessment" does not singularly mean a test or a measurement. Assessment becomes formative when it results in useful reflections and changes to the course of teaching, learning, or both.[56] The term "feedback loops" emphasizes that measurement is only part of the process. Feedback loops that lead to instructional improvement—including adaptations in teaching and learning—yield the strongest outcomes for students.

We also use "feedback loops" as a plural term because there are many types and levels of loops that are important. Students can benefit from feedback when they work individually, as a member of a small group, or in a classroom discussion. Feedback loops are valuable "in the moment"—for example, as a student practices a skill. Further, feedback loops are valuable when they cover larger spans of effort and reflections, such as at the end of presenting a project or term paper. In addition, feedback loops can assist teachers, for example, helping them notice

---

[53] Mostow, J., Aist, G., Burkhead, P., Corbett, A., Cuneo, A., Eitelman, S., Huang, C., Junker, B., Sklar, M.B., & Tobin, B. (2003). Evaluation of an automated reading tutor that listens: Comparison to human tutoring and classroom instruction. *Journal of Educational Computing Research, 29(1),* 61–117. https://doi.org/10.2190/06AX-QW99-EQ5G-RDCF

[54] Shute, V.J., Ventura, M., & Kim, Y.J. (2013). Assessment and learning of qualitative physics in Newton's Playground. *The Journal of Educational Research*, 106(6), 423–430. https://doi.org/10.1080/00220671.2013.832970

[55] Shute, V J. (2008). Focus on formative feedback. *Review of Educational Research, 78(1),* 153–189. https://doi.org/10.3102/0034654307313795

[56] Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21(1),* 5-31. https://doi.org/10.1007/s11092-008-9068-5

their own patterns of responding to students' ideas. Moreover, feedback loops are critical to the continuous improvement of products and the implementation of programs.

Due to the importance of feedback loops, formative assessment could be a leading area for schools' explorations of powerful uses of AI in teaching and learning. Educators can build upon alignments between their long-standing visions for formative assessment and the emerging capabilities that AI holds. Further, the professional assessment community brings a toolkit for asking and answering questions about topics like bias and fairness. The psychometric toolkit of methods is a strong start toward the questions that must be asked and answered because it already contains ways to measure bias and fairness and, more generally, to benchmark the quality of formative assessments. But as our discussion reveals, AI can only make feedback loops better if we keep a firm eye on the weaknesses of AI and how AI introduces new concerns.

An Example: Automated Essay Scoring

One instructive example is Automated Essay Scoring (AES). To become strong writers, which is a valuable life skill, students need regular and specific feedback. However, reviewing and providing feedback on essays is very time consuming for humans. Hence, Ellis Page provided a first vision for *computer programs that could review and provide feedback on student essays* in 1966[57], and much effort has gone into AES technologies in the intervening 56 years. Many research review articles are available to summarize the progress, which has been impressive.[58] Further, some of today's applications of AES technologies will be familiar to readers, such as Grammarly, Turnitin, and the various essay analysis engines used by publishers and assessment companies. Also note that while the traditional AES functionality emphasizes scoring or rating essays, newer AI-enabled products focus more on providing students with constructive criticism and developing their skills as writers. Writing is a life skill that is important to the pursuit of college and career ambitions, and developing writers require comprehensive feedback. If developers could inexpensively augment human feedback to developing writers with AI feedback, it's possible that support for learning to write could become more equitable.

And yet, AES is an instructive example because researchers have analyzed limitations, too.[59] AES technologies in AI can analyze some features of student essays but can also be misled by the length of an essay, by a student who places appropriate keywords in sentences that don't make sense, and other flaws that a human reader would easily notice. In a telling quote, one team that reviewed the state of the art wrote this:

The authors further note that while human and AI judgements of essays may correlate, people and computers are not noticing the same things in student writing. Due to these limitations, we must continue to emphasize a human in the loop foundation for AI-enhanced formative assessment. AI may support but not replace high-quality, human-led processes and practices of formative assessment in schools.

---

[57] Page, E.B. (1966). The imminence of grading essays by computer. *Phi Delta Kappan, 47(5),* 238–243

[58] Ke, Z., & Ng, V. (2019). Automated essay scoring: A survey of the state of the art. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence,* 6300–6308. https://doi.org/10.24963/ijcai.2019/879

[59] Doewes, A. & Pechenizkiy, M. (2021). On the limitations of human-computer agreement in automated essay scoring. In *Proceedings of the 14th International Conference on Educational Data Mining (EDM21).* https://educationaldatamining.org/EDM2021/virtual/static/pdf/EDM21_paper_243.pdf

"Nevertheless, the time when AES systems will be able to operate on a par with human judges, with similar levels of connoisseurship for such features as meaning, emotion, originality, creativity, fluency, sense of audience and so on, arguably remains a long way off."
—Gardner, O'Leary, and Yuan[60]

Key Opportunities for AI in Formative Assessment

Based on the listening sessions we held, we see three key areas of opportunity in supporting formative assessment using AI systems and models.

First, we recommend a strong focus on **measuring what matters**[61] and particularly those things that have not been easily measured before and that many constituents would like to include in feedback loops. The example above, AES, was chosen because writing remains a valuable academic, workplace, and life skill. Looking at community goals through the lens of their visions for their high school graduates, we see that families/caregivers, students, and community leaders want to nurture graduates who solve problems adaptively, who communicate and collaborate well, who persevere and self-regulate when they experience challenges. "What matters" today reaches beyond a sole focus on the core academic content measured by large-scale summative assessments, to support students and teachers with actionable feedback that nurtures the broader skills students need to succeed and thrive. Further, within core academic content, AI may help us to provide feedback on the more realistic and complex aspects of doing math, for example, or investigating scientific phenomena, understanding history, or discussing literature.

Second, we'd like to see a strong focus on **improving help-seeking and help-giving**.[62] Asking for and giving help is crucial to learning[63] and practicing a growth-mindset and central to the notion of human feedback loops. Students may not always know when they need help. In one example, computer algorithms can detect a student who is "wheel spinning" (working hard on mastering content but not making progress).[64] A student who is working hard may not feel like they need help, and the teacher may not be aware that the student is struggling if he or she appears to be "on task." AI may also be helpful by highlighting for students and teachers what forms of assistance have been most useful to the student in the recent past so that an educator can expand access to specific assistance that works for that individual student. Finally, educators may learn things from AI-enabled systems and tools that give feedback and hints during the completion of

---

[60] Gardner, J., O'Leary, M. & Yuan, L. (2021). Artificial intelligence in educational assessment: "Breakthrough? Or buncombe and ballyhoo?" *Journal of Computer Assisted Learning, 37(5),* 1207–1216. https://doi.org/10.1111/jcal.12577

[61] Merrill, S. (2020). In schools, are we measuring what matters? *Edutopia*. https://www.edutopia.org/article/schools-are-we-measuring-what-matters

[62] Roll, I., Aleven, V., McLaren, B.M., Koedinger, K.R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system, *Learning and Instruction, 21(2),* 267–280. https://doi.org/10.1016/j.learninstruc.2010.07.004.

[63] Webb, N.M., & Farivar, S. (1994). Promoting helping behavior in cooperative small groups in middle school mathematics. *American Educational Research Journal, 31(2),* 369–395. https://doi.org/10.3102/00028312031002369

[64] Kai, S., Almeda, M.V., Baker, R. S., Heffernan, C., & Heffernan, N. (2018). Decision tree modeling of wheel-spinning and productive persistence in skill builders. *Journal of Educational Data Mining,* 10(1), 36–71. https://doi.org/10.5281/zenodo.3344810

homework, utilizing that feedback to later reinforce concepts in direct instruction and strengthen the one-on-one support provided to students.[65] AI-enabled systems and tools can provide teachers with additional information about the students' recent work, so their instructor has a greater contextual sense as they begin to provide help.

Third, we advocate for **teachers and students to be strongly involved in designing feedback loops** as developers produce AI-enhanced formative assessments so they can directly voice what would make assessments less onerous and more convenient and valuable to them.[66] Earlier in the *Teaching* section, we emphasized how important it is to involve teachers in designing, selecting, and evaluating AI-enhanced technologies. Students need to be centered, too. They are experiencing AI in their everyday lives, and they have strong opinions on what is valuable and safe. There are local and cultural variations in how people provide and receive feedback, so adjusting feedback to align with community norms is important.

Key Recommendation: Harness Assessment Expertise to Reduce Bias

Bias and fairness are important issues in assessment design and administration,[67] and they hold relevance for the area of AI-enabled assessment. In traditional assessment, a test item might be biased if unnecessary details are included that differentially advantage some students (e.g., a story-based item that references a sport that only boys play regularly may be less helpful to girls). As discussed earlier, with AI, we now must worry about algorithmic discrimination which can arise due to the manner in which AI algorithms are developed and improved from large datasets of parameters and values that may not represent all cohorts of learners.

Algorithmic discrimination is not just about the measurement side of formative assessment; it is also about the feedback loop and the instructional interventions and supports that may be undertaken in response to data collected by formative assessments. There is a question both about access to such interventions and the quality or appropriateness of such interventions or supports. When an algorithm suggests hints, next steps, or resources to a student, we have to check whether the help-giving is unfair because one group systematically does not get useful help which is discriminatory. Fairness goes beyond bias as well. In AI-enabled formative assessment, both the opportunity to learn through feedback loops, as well as the quality of learning in and outside of such loops, should be addressed. Issues of bias and fairness have arisen in traditional assessments, and the field of psychometrics has already developed valuable tools to challenge and address these issues.[68] Assessment as a field may have a head start on tackling bias and fairness for AI in education. And yet the issues expand with AI, so the work is not done. Strong and deliberate attention to bias and fairness is needed as future formative assessments are developed.

---

[65] Walker, E., Rummel, N. & Koedinger, K.R. (2015). Adaptive intelligent support to improve peer tutoring in algebra. *International Journal of Artificial Intelligence in Education, 24*, 33–61 https://doi.org/10.1007/s40593-013-0001-9

[66] Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J.M., Milligan, S., Selwyn, B. & Gašević,D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence, 3*. k https://doi.org/10.1016/j.caeai.2022.100075

[67] Reynolds, C.R., & Suzuki, L.A. (2012). Bias in psychological assessment: An empirical review and recommendations. *Handbook of Psychology, Second Edition*. https://doi.org/10.1002/9781118133880.hop210004

[68] Kaplan, R.M., & Saccuzzo, D.P. (2017). *Psychological testing: Principles, applications, and issues*. Cengage Learning.

Related Questions

As indicated, formative assessment is an area in which AI is expanding along a continuum that can be guided by visions already in place, such as the 2017 NETP. It is an area in which AI is poised to grow, especially with capabilities that power more feedback loops in student learning. As this growth takes place, we suggest ongoing attention to the following questions:

- Is formative assessment bringing benefits to the student learning experience and to the efficacy of classroom instruction?
- Are humans being centered in AI-enabled formative assessment and feedback loops?
- Are we providing empowering professional development to teachers so they can leverage feedback loops and safeguard against concerns?
- To what extent are the developers and implementers of AI-enabled systems and tools tackling new sources of algorithmic bias and continuing to make assessment fairer?
- Are governance policies regarding who owns, controls, and can view or use AI-enabled formative assessment data appropriate and adequate?
- Do we have sufficient guardrails against misuse of formative assessment data or automatically generated interpretations of student achievement and learning, such as on dashboards?
- Is trust in an AI-enabled assessment system, feedback loops, and data generated by such assessments growing or diminishing?

Research and Development

Policy relies upon research-based knowledge; likewise, improving practice depends on feedback loops that analyze empirical evidence. Consequently, the 2010 NETP specified a series of "grand challenges" which were "R&D problems that might be funded and coordinated at a national level." One 2010 NETP grand challenge was to create personalized learning systems that continuously improve as they are used:

> *"Design and validate an integrated system that provides real-time access to learning experiences tuned to the levels of difficulty and assistance that optimize learning for all learners and that incorporates self-improving features that enable it to become increasingly effective through interaction with learners."*[69]

Since 2010, much R&D has addressed this challenge. Conferences about learning analytics, educational data mining, and learning at scale have blossomed. Developers have created platforms that use algorithms and the analysis of big data to tune learning experiences. The challenge has not been fully achieved, and further work on this challenge is still relevant today.

Insight: Research Can Strengthen the Role of Context in AI

Despite the relevance of 2010's grand challenges, it has become apparent that the R&D community is now looking to expand their attention. The 2010 challenges were stated as technical problems. Today's researchers want to more deeply investigate context, and today's tech companies want to develop platforms that are responsive to the learners' characteristics and situations more broadly—not just in terms of narrow cognitive attributes. We see a push to transform R&D to address context sensitivity. We look forward to new meanings of "adaptive" that broaden outward from what the term has meant in the past decade. For example, "adaptive" should not always be a synonym of "individualized" because people are social learners. Researchers therefore are broadening "adaptivity" to include support for what students do as they learn in groups, a form of learning that is prevalent in schools across the U.S.

The focus on context is not an accident. Context is a traditional challenge in AI.[70] Thus, researchers and developers are wise to prioritizing context. Unless we invest more in AI that is context-sensitive, it is quite likely that AI will break and fail to achieve educational goals.

Agreeing to prioritize context won't be easy. As illustrated above in Figure 12, there will be a tension between depth of context and pace of technological advances in AI R&D. On the one hand, AI is sometimes presented as a race to be the first to advance new techniques or scale new applications—innovation is sometimes portrayed as rapidly going to scale with a minimally viable product, failing fast, and only after failure, dealing with context. On the other hand, researchers and developers see that achieving good innovations with AI in education will clearly

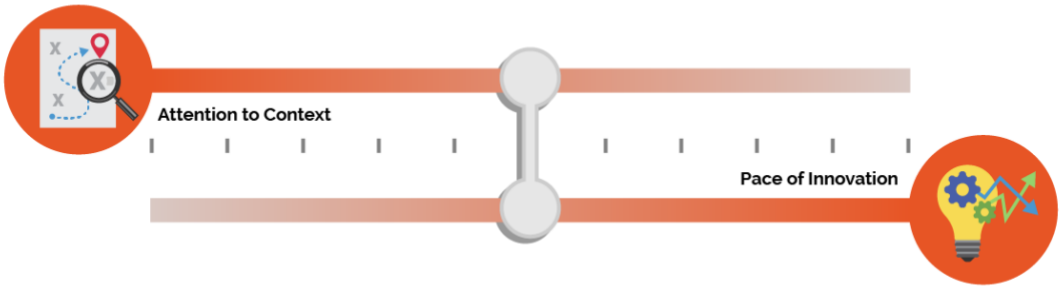
---

[69] U.S. Department of Education, Office of Educational Technology. (2010). *Transforming American Education: Learning Powered by Technology.* U.S. Department of Education. p. 78

[70] Boden, M.A. (2018). *Artificial intelligence: A very short introduction*. Oxford. ISBN: 978-0199602919

require bringing more context into the process early and often. For example, researchers highlight that humans must be continually adjusting the goals for technology and have noted that when we set forth goals, we often don't yet fully understand context; and as we learn about context, the goals must change.[71] This suggests that context must be prioritized early and habitually in R&D; we don't want to win a race to the wrong finish line.

Figure 12: The tension between depth of context and pace of technological advances in AI



Further, intensifying focus on context in this work will change the nature of the R&D. There won't be just one type of change in R&D because context has multiple meanings. Attendees in our listening sessions described four types of context necessary for the future.

We list these four types of context below and then expand on each one in its own section. These four types emerged as topics of provocations to think differently about R&D but certainly do not exhaust the important ways of investigating context.

1. **Focus on the Long Tail:** How could we use big data and AI to pay more attention to the "long tail" of edtech use—going beyond a few "most typical" ways of using emerging technology and instead solving for digital equity and inclusion?
2. **Partnership in Design-Based Research:** How can we change who is involved and influential in designing the future of AI in education to more centrally include students, teachers, and other educational constituents?
3. **Connect with Public Policy:** How can work on AI in education build on general advances in AI ethics, safety, and regulation and contribute additional advances specific to educational policy?
4. **Rethink Teacher Professional Development:** How can we solve for new systems of teacher professional development (both pre-service and in-service) that align to the increasingly core role of technology in the teaching profession?

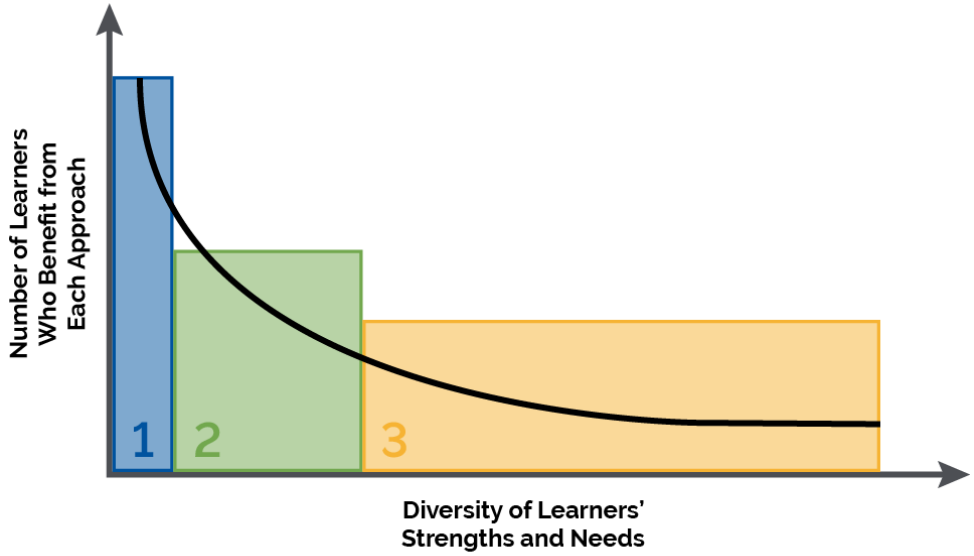
[71] Russell, S. (2019). Human compatible: Artificial intelligence and the problem of control. Penguin. ISBN: 9780525558637

"We can't necessarily always apply traditional research methodologies to this topic because educational technology changes so quickly."
—Kristina Ishmael, Office of Educational Technology

Attention to the Long Tail of Learner Variability

At the core of R&D of AI in education, innovators will be building models that fit available data. The increasing scale and prevalence of technologies means that the data is coming from and including a wide range of different contexts and varied ways that people in those contexts engage in teaching and learning. Researchers in our listening sessions drew attention to the promise of AI for addressing "context" by reference to the long tail of learner variability.

Figure 13: The long tail of learner variability



As depicted in Figure 13, learners vary in their strengths and needs. The most frequently occurring mix of strength and needs (also known as "teaching to the middle") is depicted leftmost, with less frequently occurring mixes spreading to the right. Rising upward, the figure depicts the number of learners who benefit from a particular learning design, pathway, or approach. We argue that AI can bring opportunities to address a wider spectrum of strengths and needs but only if developers and innovators focus on the long tail and not only "teaching to the middle."

For the sake of argument, the figure indicates three zones. In a first zone, curricular resources are mostly standardized, with perhaps a dimension or two of adaptivity. For example, many existing products adapt based on the correctness of student answers and may also provide options to read or hear text in a second language. However, the core of the instructional approach is highly standardized. In a second zone, there is greater balance between how much standardization and how much adaptivity students can access. Universal Design for Learning (UDL) is one set of recommendations for providing learning opportunities in multiple formats and for

accommodating different learning progressions.[72] UDL can enable accommodating more ways in which learners vary, and as teachers know, there are many more important ways to adapt to students than found in today's edtech products.

Students are neurodiverse. They bring different assets from their experiences at home, in their communities, and in their cultures. They have different interests and motivations. And they learn in varied settings—classrooms and schools differ, and at-home students learn in informal settings in ways that could complement school learning. These are all important dimensions of "context." Zone 3 indicates highly adaptive learning, where standardization is less successful and where we need to discover a wider variety of approaches to engage learners and sustain powerful learning. Researchers in our listening sessions noted the promise of Zone 3 because AI's ability to recognize patterns in data can extend beyond the most common patterns and because AI's ability to generate customized content can extend beyond what people can reasonably generate on their own.

Notice that although the Zone 1 bar appears to be the tallest, and thus tends to attract initial attention, there are more students in Zones 2 and 3, the regions where AI can provide more help. Thus, it's important to ask where AI researchers and developers are directing their attention. When we say a model "fits," are we saying it fits the most common and typical uses by teachers and learners? This sort of R&D is easier to do. However, machine learning and AI also can tailor a model to the less common and more culturally specific contexts, too. Therefore, how can constituents cultivate interdisciplinary expertise to direct attention among researchers and developers to focus on the long tail? If we do, the quality of what we do for those represented in that tail can be more adaptive and more context-sensitive. And to be most effective, it will require the integration of contextual, content, and technical expertise.

Within the long-tail challenge, the community is wondering how we can get to research insights that are both general and specific enough. When research produces very general abstractions about learning, it often doesn't give developers enough guidance on exactly how to adjust their learning environments. Conversely, when research produces a specific adaptive algorithm that works on one educational platform, it often remains hard to apply to additional platforms; research can be too detailed as well. The research community is also thinking about new partnerships that could bring more data and more diverse perspectives to the table, the topic of the next section.

Focusing on the long tail of learner variability is particularly important to addressing a long-standing key research question: *"Do new AI-enhanced approaches work to improve learning, and for whom and under what conditions?"*

Partnership in Design-Based Research

Of course, teachers must be included in rethinking their own professional development. This thought leads to another priority aspect of context: partnership in design-based research. With regard to inclusive design, attendees in our listening sessions brought up a variety of co-design[73]

---

[72] Rose, D. (2000). Universal design for learning. *Journal of Special Education Technology*, *15*(4), 47-51. https://doi.org/10.1177/016264340001500407

[73] Roschelle, J., Penuel, W., & Shechtman, N. (2006). Co-design of innovations with teachers: definition and dynamics. In *Proceedings of the 7th International Conference on Learning Sciences*, Bloomington, IN. https://doi.dx.org/10.22318/icls2006.606

and other participatory processes and goals that can be used in R&D.[74] By co-design, they mean sharing power with non-researchers and non-developers through all the phases of design and development, which would result in more influence by teachers, students, and other constituents in the shape of AI-enabled edtech. The shift toward co-design was palpable throughout our listening sessions, but as researchers and developers have not standardized on one particular co-design method, we share some representative examples.

- Youth can powerfully participate in design when researcher methods include participant co-design. Such research can investigate how to improve edtech while educating students. A listening session attendee asked about developing students' awareness of what data are being collected and how data are being used by developers.
- There is a near future need to go beyond representation so that co-designed solutions consider more generous contexts for broader possibilities, according to attendees.
- The shift of power dynamics is another research-worthy interest of the panel and attendees to understand the balance between a teacher's agency and a machine's suggestions.
- Likewise, such longitudinal research will require both the infrastructure and institutional support to fund necessary experimentation and requisite failures to elicit positive results and safe innovation.
- There is a desire for rapid cycle evaluations with inclusive feedback loops that return to the educators themselves as essential relative to traditional research approaches.
- Many researchers also mentioned a focus on explainable AI as essential to enable participation in the design and evaluation of emerging AI approaches in education.

The conversations raised this question: how can co-design provide an empowering form of participation in design and thus achieve digital inclusion goals? Such digital inclusion can span many layers of design, including diverse representation in design of policies around data, design of adaptivity, and other user experiences in AI systems, design of plans for cultivating AI literacy for users of new platforms, and lastly, the design of plans to evaluate systems.

Re-thinking Teacher Professional Development

With regard to teachers as professionals, both researchers and other educators attending our listening sessions were highly concerned about the disconnect between how teachers are prepared versus how they are expected to work with emerging technology. When we discuss learning, teachers are central actors, and thus the contexts in which they are prepared is centrally important to their ability to do great work in current and emerging technological environments.

Teacher professional development, professional learning, and leadership (PD or PL) for emerging technologies was seen as an area needing intense re-thinking, and research could lead the way. Today, few who prepare to become a teacher in an established pre-service program learn about the effective use of educational technology in schools and classrooms; those who do

[74] Center for Integrative Research in Computing and Learning Sciences (CIRCLS). (2022, Feb.). From Broadening to Empowering: Reflecting on the CIRCLS'21 Convening. https://circls.org/circls21report

have the opportunity to investigate technology rarely think about the structures that shape its use in the classroom and in educational leadership. Consequently, a troubling dichotomy arises between a small set of investigators who specifically consider educational technology in their research on teaching and a broader group of educators who see educational technology as a generic instructional resource. The challenge is high because teacher professional development will remain highly varied by local contexts. Yet insufficient attention to teachers as leaders in the use and further development of effective educational technology is widespread in teacher professional development research.

One response can be in terms of investigating how to nurture greater AI literacy for all teachers. AI literacy is not only important to protect educators and students from possible dangers but also valuable to support teachers to harness the good and do so in innovative ways. A panelist reminded the group that this work implies how we prepare educators with a baseline AI literacy and understanding. More transparency and authentic dialogue can foster trust, which was mentioned by a researcher as a chief concern for all teachers and students.

This is not to suggest that AI literacy is a complete or even a simple fix. Researchers want to ask fundamental questions about what it means for teachers to be professionals, especially as emerging technologies gain ground in schools and classrooms—our teachers' professional workplaces. Researchers want to broadly reconceptualize teacher professionalism and to stop treating technology as an add-on element of professional development.

Connecting with Public Policy

Defining human-centered AI for education requires the embrace of a human-centered principle and foundation for developing and formulating policies that govern the application and use of AI more generally throughout society. For example, power dynamics that arise between companies and consumers in society around issues like data ownership will also arise in the education-specific ecosystem. Further, the public discourse in which people are discussing ethics, bias, responsibility, and many other necessary concepts will be happening simultaneously in public policy and in educational ecosystems.

One clear implication in our listening sessions was that efforts to improve AI literacy in education could be important and helpful to society more generally. For example, one panelist said that an overarching goal of improving AI literacy is necessary if they are to contribute to how those technologies are designed. Another researcher was interested in how edtech can provide environments where students can experience having difficult discussions across perspectives, an issue which is endemic to present society. A third researcher noted the insufficiencies of prior efforts to contend with algorithmic bias, ethics, and inclusion due to a classroom's complex social dynamics.

Researchers want to take a lead in going beyond checkbox approaches to take these issues seriously. And they also acknowledge that engaging with policy is often a new form of context for edtech and AI researchers, many of whom don't have long experiences in policy arenas. Likewise, developers often do have experience with some policy issues, such as data privacy and security, but are now needing to become part of new conversations about ethics, bias,

transparency, and more, a problem that the EdSAFE AI Alliance is addressing through multi-sector working groups and policy advocacy.[75]

Key Recommendation: Focus R&D on Addressing Context

Attendees who have participated in listening sessions leading up to this report were exceptionally clear that their view of future R&D involved a shift from narrow technical questions to richer contextual questions. This expansive shift toward context, as detailed below, is the foundational orientation that the listening session attendees saw as being necessary to advancing R&D. Attendees included these as dimensions of context:

- learner variability, e.g., in disabilities, languages spoken, and other relevant characteristics;
- interactions with peers, teachers, and others in the learning settings;
- relationships across home, school, and community settings, including cultural assets;
- instructional resources available while learning;
- teacher preparation; and
- policies and systems that structure teaching and learning.

To more fully represent the context of teaching and learning, including these and other dimensions of text, researchers will have to work in partnership with others to understand which aspects of context are most relevant to teaching and learning and how they can be usefully incorporated into AI models.

Ongoing Questions for Researchers

As mentioned earlier, people are good at context; AI—not so much. R&D investment in context-rich edtech thus could serve multiple national interests because finding ways to do a better job with context would be a fundamental advancement in AI. Indeed, questions like these reverberate across all applications of AI in society, and education is a centrally good context for investigating them:

- Are AI systems moving beyond the tall portions of the "long tail" to adapt to a greater range of conditions, factors, and variations in how people learn?
- To what extent are AI technologies enhancing rather than replacing human control and judgment of student learning?
- How will users understand the legal and ethical implications of sharing data with AI enabled technologies and how to mitigate privacy risks?
- To what extent does technology account for the complex social dynamics of how people work and learn together, or is technology leading humans to narrow or oversimplify?
- How can we more clearly define what we mean by a context-sensitive technology in terms that are both concrete and broad enough? How can we measure it?

---

[75] Nentrup, E. (2022). How Policymakers Can Support Educators and Technology Vendors Towards SAFE AI. EdSAFE AI Alliance. https://www.edsafeai.org/post/how-policymakers-can-support-aied

- To what extent are technical indicators and human observations of bias or unfairness working together with human observations? How can concerns about ethics and equity in AI technologies become actionable both in R&D, and later, when AI is widely used?
- Are we learning for whom and under what conditions AI systems produce desired benefits and impacts and avoid undesirable discrimination, bias, or negative outcomes?

Desired National R&D Objectives

Attendees sought immediate progress on some key R&D issues, such as these:

- Clarifying and achieving a consensus on the terms that go beyond data privacy and data security, including ideas like human-centered, value-sensitive, responsible, ethical, and safe so constituents can advocate for their needs meaningfully and consistently
- Creating and studying effective programs for AI literacy for students, teachers, and educational constituents in general, including literacy with regard to the ethics and equity issues specific to AI in educational settings
- Advancing research and development to increase fairness, accountability, transparency, and safety in AI systems used in educational settings
- Defining participatory or co-designed research processes that include educators in the development and conduct of research related to the development, use, and efficacy of AI-enabled systems and tools
- Highlighting and advancing R&D efforts that empower the participation and voices of youth regarding research, data, and design of AI applications for teaching and learning

Longer term desires for a national R&D program include some of the following objectives:

- Funding sustainable partnerships that uncover what context means and how it can be addressed over longer periods of time
- Better connecting goals for "broadening participation" (for example, in STEM learning pathways) to strategies for addressing learner variability and diversity
- Prioritizing research to revitalize support for instructors in light of the increasingly technological nature of K-12, higher education, and workplace learning settings
- Creating infrastructure and new ways of working together beyond individual field-initiated grants so that R&D with big data and leveraging emerging AI capabilities becomes safer and more productive

Recommendations

Earlier, we asked two guiding questions:

1. What is our collective vision of a desirable and achievable educational system that leverages automation while protecting and centering human agency?

2. On what timeline will we be ready with necessary guidelines and guardrails along with convincing evidence of positive impacts, so that we can ethically and equitably implement this vision widely?

Answers to the first question are provided throughout the *Learning*, *Teaching*, *Assessment*, and *Research* sections. This section turns to a call to action to education leaders and to recommendations. Core to the Department's perspective is that education will need leadership specific to our sector. Leadership should recognize and build on prior accomplishments in edtech (such as strong prior work on student privacy and school data security) as well as broad frameworks for safe AI (such as the *Blueprint for an AI Bill of Rights*). Leadership must also reach beyond these accomplishments and frameworks to address emerging opportunities and risks that are specific to novel capabilities and uses of AI in education.

Insight: Aligning AI to Policy Objectives

Individual sections of this policy report provided insights in each of four areas—learning, teaching, assessment, and research. These insights, synthesized from extensive stakeholder consultation and listening sessions, show that the advances in AI can bring opportunities to advance the Department's policy objectives:

- In support of our objective of attracting and retaining teachers, our nation could focus on AI assistants that make teaching jobs better and provide teachers with the information they need to work closely and empathically with students. An emphasis on teachers in the loop could ensure that AI-enabled classroom technologies keep teachers in the know, in touch with their students, and in control of important instructional decisions. Keeping the teacher in the loop is important to managing risks, as well.

- In support of equitable learning, especially for those most affected by the pandemic, AI could shift edtech from a current deficit-based model to a strengths-based alternative. In addition to finding student weaknesses and assigning fixes, edtech could make recommendations based on strengths that students bring to learning and how adapting to the whole student—a cognitive, social, and self-regulating person—could enable more powerful learning. Adapting to the whole student should include supporting students with disabilities as well as English learners. With regard to equity, we must remain highly attuned to the challenges of bias (which are inherent to how AI systems are developed) and take firm action to ensure fairness.

- With regard to growth trajectories to successful careers, AI-enabled assessments could provide students and teachers with formative guidance on a wider range of valuable skills, focusing on providing information that enhances learning. Aligned with the human-centric view, we should take a systems view of assessments where students, teachers, and others remain at the center of instructional decision making.

- With regard to equity, as research advances and brings more context into AI, we will be better able to use AI to support goals that require customization of learning resources, such as enabling teachers to more easily transform materials to support neurodiverse learners and increase responsiveness to local communities and cultures.

Going forward, educational leaders need to bring these and their own policy priorities to the table at every discussion about AI, driving the conversation around human priorities and not only their excitement about what new technology might do. Fundamentally, AI seeks to automate processes that achieve goals, and yet, AI should never set goals. The goals must come from educators' vision of teaching and learning and educators' understanding of students' strengths and needs.

Calling Education Leaders to Action

We summarize seven recommendations for policy action. These recommendations are for education leaders. In the introduction, we note the necessity of involving education constituents in determining policies for AI. We also observed throughout our listening sessions that people coming from many different roles in education all have passion, knowledge, and insights to contribute. In our view, all types of constituents can be education leaders. We are reluctant to suggest any constituent role is more important to advance any of the recommendations, but we call out specific needs for action within some of the recommendations where it is warranted.

Recommendation #1: Emphasize Humans in the Loop

We start with a central recommendation throughout this report. This recommendation was a clear constituent favorite. Indeed, across more than 700 attendees in our listening sessions, the predominant discussion tackled how constituents can achieve a consensus vision for AI-enabled edtech where humans are firmly at the center. The *Blueprint for an AI Bill of Rights* similarly calls for "access to timely human consideration and remedy by a fallback and escalation process if an automated system fails, it produces an error, or you would like to appeal or contest its impacts…" Building on this consensus, we call upon all constituents to adopt "humans in the loop" as a key criterion for educational use of AI.

We envision a technology-enhanced future more like an electric bike and less like robot vacuums. On an electric bike, the human is fully aware and fully in control, but their burden is less, and their effort is multiplied by a complementary technological enhancement. Robot vacuums do their job, freeing the human from involvement or oversight.

Although teachers should not be the only humans involved in loops, Figure 5 provided examples of three types of teacher loops that are central to education and can be used to illustrate what "human in the loop" means. Here, we use the example of an AI chatbot to elaborate on the meaning of the loops. First, as students become involved in extended interactions with AI chatbots, teachers will need to educate students about safe AI use, monitor their use, and provide human recourse when things go astray. Second, teachers are beginning to use chatbots to plan personalized instruction for their students; they will need to be involved in loops with other teachers to understand effective prompts, to know how to analyze AI-generated lesson plans for flaws, and to avoid the human tendency to overly trust AI systems and underapply human judgement. Third, teachers need to be involved in the design and evaluation of AI systems before they are used in classrooms and when needs for improvement are observed. In one example, to design AI-generated homework support for students, teachers' in-depth understanding of the

cognitive, motivational, and social supports their students need will provide much-needed guidance as a homework-support chatbot is designed.

In framing AI in education, this report advances a key recommendation of "human in the loop" AI because the phrase readily communicates a criterion that everyone can use as they determine which AI-enabled systems and tools are appropriate for use in teaching and learning. In a rather technical field, human in the loop is an approachable and humanistic criterion. Rather than suggesting that AI-enabled systems and tools should replace teachers, this term instead solidifies the central role of educators as instructors and instructional decision makers, while reinforcing the responsibility of teachers to exercise judgement and control over the use of AI in education. It resonates with the important idea of feedback loops, which are highly important to how people teach and learn. It also aligns with the ideas of inspectable, explainable, severable, and overridable AI.

The Department agrees with listening session participants who argued that teachers should not be the only humans in the loop and calls upon parents, families, students, policy makers, and system leaders to likewise examine the "loops" for which they are responsible, critically analyze the increasing role of AI in those loops, and determine what they need to do to retain support for the primacy of human judgement in educational systems.

Recommendation #2: Align AI Models to a Shared Vision for Education

> *"All models are wrong, but some are useful."*
> *—George Box, Statistician*

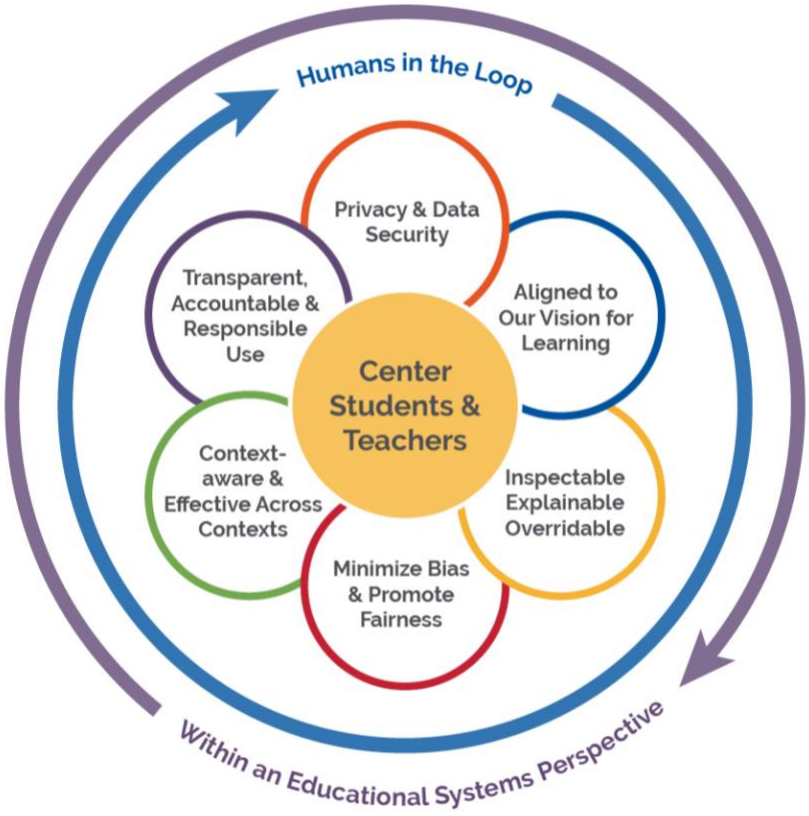
As we have discussed across every section of this report, AI technologies are grounded in models, and these models are inevitably incomplete in some way. It is up to humans to name educational goals and measure the degree to which models fit and are useful—or don't fit and might be harmful. Such an assessment of how well certain tools serve educational priorities may seem obvious, but the romance of technology can lead to a "let's see what the tech can do" attitude, which can weaken the focus on goals and cause us to adopt models that fit our priorities poorly.

Here we call upon educational policy and decision makers at the local, state, and federal level to use their power to align priorities, educational strategies, and technology adoption decisions to place the educational needs of students ahead of the excitement about emerging AI capabilities. We want to strengthen their attention to existing state, district, and school-level policies that guide edtech adoption and use, such as the four levels of evidence in ESSA, the privacy requirements of FERPA, and enhanced policies to come. Local education leaders know best what their urgent educational priorities are. Every conversation about AI (or any emerging technology) should start with the educational needs and priorities of students front and center and conclude with a discussion about the evaluation of effectiveness re-centered on those needs and priorities. Equity, of course, is one of those priorities that requires constant attention, especially given the worrisome consequences of potentially biased AI models.

We especially call upon leaders to avoid romancing the magic of AI or only focusing on promising applications or outcomes, but instead to interrogate with a critical eye how AI-enabled systems and tools function in the educational environment. We ask leaders to distrust broad claims and ask six types of questions, listed below. Throughout this report, we elaborated on

which characteristics of AI model use in education are most important to evaluate for alignment to intended educational goals. To aid leaders, we summarize our insights about AI models and their use in educational tools and systems in Figure 14.

Figure 14: Recommendation for desired qualities of AI tools and systems in education



In this figure, we center teaching and learning in all considerations about the suitability of an AI model for an educational use. Humans remain in the loop of defining, refining, and using AI models. We highlight the six desirable characteristics of AI models for education (elaborating from principles in the *Blueprint for an AI Bill of Rights* to fit the specifics of educational systems):

1. **Alignment of the AI Model to Educators' Vision for Learning**: When choosing to use AI in educational systems, decision makers prioritize educational goals, the fit to all we know about how people learn, and alignment to evidence-based best practices in education.

2. **Data Privacy:** Ensuring security and privacy of student, teacher, and other human data in AI systems is essential.

3. **Notice and Explanation:** Educators can inspect edtech to determine whether and how AI is being incorporated within edtech systems. Educators' push for AI models can explain the basis for detecting patterns and/or for making recommendations, and people retain control over these suggestions.

4. **Algorithmic Discrimination Protections:** Developers and implementers of AI in education take strong steps to minimizing bias and promoting fairness in AI models.

5. **Safe and Effective Systems:** The use of AI models in education is based on evidence of efficacy (using standards already established in education for this purpose) and work for diverse learners and in varied educational settings.

6. **Human Alternatives, Consideration and Feedback:** AI models that support transparent, accountable, and responsible use of AI in education by involving humans in the loop to ensure that educational values and principles are prioritized.

Although we first address our recommendation to interrogate how educational systems use AI models to educational leaders who adopt technologies, other leaders also have integral roles to play. Teachers and students, as well as their families/caregivers, contribute significantly to adoption decisions also. And leaders and parents must support educators when they question or override an AI model based on their professional wisdom. Developers of technologies need to be forthcoming about the models they use, and we may need policymakers to create requirements for disclosure so that the marketplace can function on the basis of information about AI models and not only by the claims of their benefits.

We also emphasize the need for a government role. AI models are made by people and are only an approximation to reality. Thus, we need policies that require transparency about the AI models that are embedded in educational systems, as well as models that are inspectable, explainable, and overridable. Our listening sessions featured constituent calls for government doing more to hold developers accountable for disclosing the types of AI models they employ in large-scale products and the safeguards included in their systems. Government leaders can make a positive contribution to market conditions that enable building trust as AI systems are procured and implemented in education. We discuss these guidelines more in recommendation #4, which is about building trust.

Recommendation #3: Design Using Modern Learning Principles

We call for the R&D sector to ensure that product designs are based on best and most current principles of teaching and learning. The first decade of adaptivity in edtech drew upon many important principles, for example, around how to sequence learning experiences and how to give students feedback. And yet the underlying conception was often deficit-based. The system focused on what was wrong with the student and chose pre-existing learning resources that might fix that weakness. Going forward, we must harness AI's ability to sense and build upon learner strengths. Likewise, the past decade of approaches was individualistic, and yet we know that humans are fundamentally social and that learning is powerfully social. Going forward, we must build on AI capabilities that connect with principles of collaborative and social learning and which respect the student not just for their cognition but also for the whole human skill set. Going forward, we also must seek to create AI systems that are culturally responsive and culturally sustaining, leveraging the growth of published techniques for doing so. Further, most early AI systems had few specific supports for students with disabilities and English learners. Going forward, we must ensure that AI-enabled learning resources are intentionally inclusive of these students. The field has yet to develop edtech that builds upon each student's ability to make choices and to self-regulate in increasingly complex environments. We have to develop edtech that expands students' abilities to learn in creative modes and to expand their ability to discuss, write, present, and lead.

We also call upon educators to reject uses of AI that are based solely on machine learning from data—without triangulation based on learning theory and knowledge from practice. Achieving

effective and equitable educational systems requires more than processing "big data," and although we want to harness insights from data, human interpretation of data remains highly important. We reject a technological determinism in which patterns in data, on their own, tell us what to do. Applications of AI in education must be grounded in established, modern learning principles, the wisdom of educational practitioners, and should leverage the expertise in the educational assessment community around detecting bias and improving fairness.

Recommendation #4: Prioritize Strengthening Trust

Technology can only help us to achieve educational objectives when we trust it. Yet, our listening sessions revealed the ways in which distrust of edtech and AI is commonplace. Constituents distrust emerging technologies for multiple reasons. They may have experienced privacy violations. The user experience may be more burdensome than anticipated. Promised increases in student learning may not be backed by efficacy research. They may have experienced unanticipated consequences. Unexpected costs may arise. Constituents may distrust complexity. Trust needs to incorporate safety, usability, and efficacy.

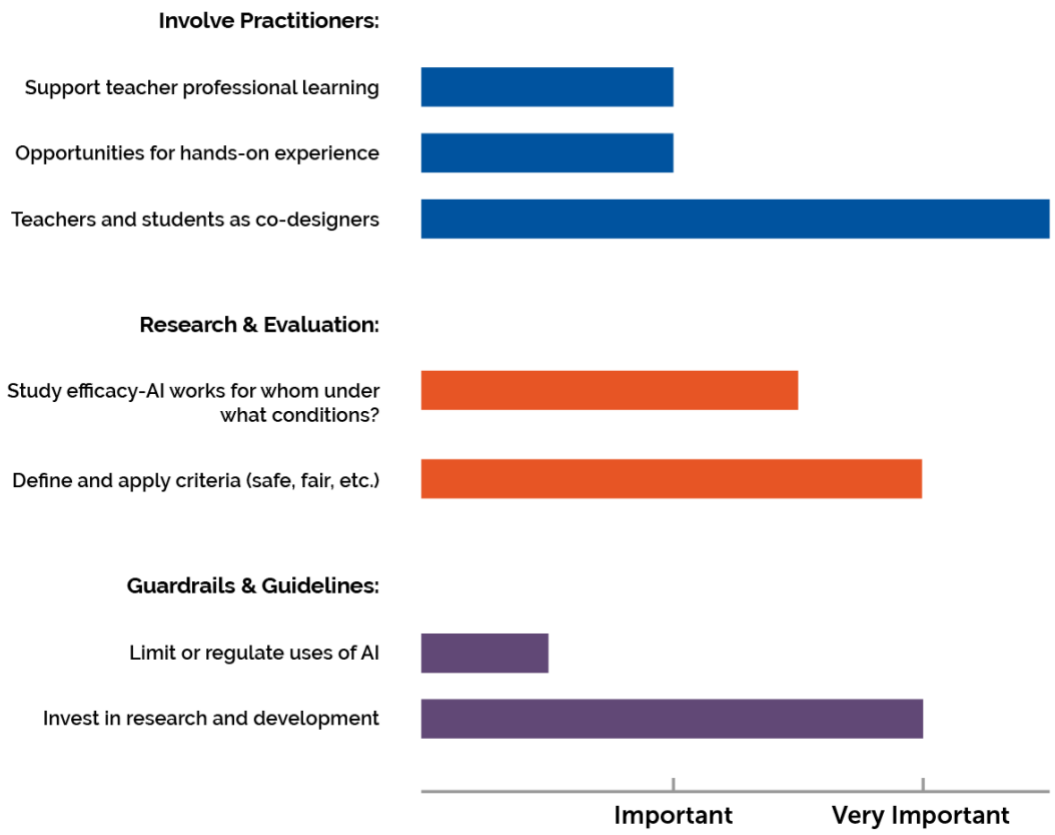
The Department firmly takes the stance that constituents want AI that supports teachers and rejects AI visions that replace teachers. And yet, teachers, students, and their families/caregivers need support to build appropriate levels of trust in systems that affect their work. In the broader ecosystem, trustworthy AI is recognized as a multidimensional problem (including the dimensions of Figure 14, above). If every step forward does not include strong elements of trust building, we worry that distrust will distract from innovation serving the public good that AI could help realize.

We expect that associations and societies have a key role in strengthening trust. Some important associations like the State Educational Technology Directors Association and the Consortium for School Network work with edtech leaders, and parallel organizations like EDUCAUSE work with postsecondary leaders. Other associations and societies work with teachers, education leaders, and education staff developers. Industry networks, like the EdSAFE AI Alliance, can bring together industry leaders to work together to foster trust. Additional societies bring researchers together. These societies and associations have the reach necessary to bring all parts of the educational ecosystem into discussions about trust and also the ability to represent the views of their constituents in cross-cutting policy discussions.

Recommendation #5: Inform and Involve Educators

Our listening sessions also asked for more specific direction on the question of what education leaders should do (see Figure 15). The most frequent responses fit three clusters: the need for guidelines and guardrails, strengthening the role of teachers, and re-focusing research and development. These are activities that constituents are asking for and that could expand trust. The recommendations that follow respond to these requests.

Figure 15: Listening session attendees prioritized involving practitioners, research, and evaluation and the need for guidelines and guardrails.



In particular, one concern that repeatedly arose in our listening sessions was the potential for AI to result in less respect for educators or less value for their skills. Across the nation, we are now responding to decreasing interest in entering or remaining in the teaching profession. Now is the time to show the respect and value we hold for educators by informing and involving them in every step of the process of designing, developing, testing, improving, adopting, and managing AI-enabled edtech. This includes involving educators in reviewing existing AI-enabled systems, tools, and data use in schools, designing new applications of AI based on teacher input, carrying out pilot evaluations of proposed new instructional tools, collaborating with developers to increase the trustworthiness of the deployed system, and raising issues about risks and unexpected consequences as the system is implemented.

We have already seen educators rise to the challenge of creating overall guidelines, designing specific uses of available AI-enabled systems and tools, and ferreting out concerns. And yet, the influence of educators in the future of AI-enabled products cannot be assumed; instead, constituents need policies that put muscle behind it. Could we create a national corps of leading educators representing every state and region to provide leadership? Could we commit to developing necessary professional development supports? Can we find ways to compensate educators so they can be at the forefront of designing the future of education? Our policies should enable educators to be closely involved in design of AI-enabled educational systems.

Although we know that the responsibility for informing and involving educators must be distributed at all levels of national and school governance, the Office of Educational Technology

can play a key role in informing and involving educators through its reports, events, outreach, and in a future NETP. Although examples above refer to K-12 teachers, higher education instructors must also be included. We also call on the edtech industry to involve educators throughout their design and development processes. For example, AI-enabled teaching assistants are only likely to help teachers do their job if teachers are thoroughly involved as the assistants are designed. We call upon institutions that prepare teachers to integrate technology more systematically into their programs; for example, the use of technology in teaching and learning should be a core theme across teacher preparation programs, not an issue that arises only in one course.

Recommendation #6: Focus R&D on Addressing Context and Enhancing Trust and Safety

Research that focuses on how AI-enabled systems can adapt to context (including variability among learners) in instructional approaches and across educational settings is essential to answering the question of, "Do specific applications of AI work in education, and if so, *for whom and under what conditions*?" The italicized phrase points to variability among learners and diversity in the settings for learning. We call upon innovators in R&D to focus their efforts to advance AI on the long tail of learning variability, where large populations of students would benefit from customization of learning. We also call on R&D to lead by establishing how trust can be strengthened in AI-enabled systems, building on the *Blueprint's* call for safe and effective systems yet also including education-specific requirements, such as how teachers can be meaningfully involved in design phases, not only in implementation and evaluation.

Although many products today are adaptive, some adapt on just one or a few dimensions of variability, such as student's accuracy in problem solving. As teachers know, there are many more important ways to adapt to students' strengths and needs. Students are neurodiverse and may have specific disabilities. They bring different assets from their experiences at home, in communities, and in their cultures. They have different interests and motivations. They are in different places in their journeys to master the English language. And they learn in varied settings. Classrooms and schools are different, and at home, students learn in informal settings in ways that could complement school learning. We recommend attention to "context" as a means for expressing the multiple dimensions that must be considered when elaborating the phrase "for whom and under what conditions." We also acknowledge the role of researchers in conducting evaluations, which must now consider not only efficacy but must also explore where harm may arise and the system problems that can occur through weak trust or over-trust in AI systems.

R&D must take the lead in making AI models more context-sensitive and ensuring that they are effective, safe, and trustworthy for use with varied learners in diverse settings. Although AI has capabilities to find patterns beyond the limited number of variables that people normally think about, AI is not particularly good at understanding and working with context in the ways people do. Over time, we've seen learning sciences grow to be less about individualistic cognitive principles and more encompassing first of social learning and then of the many dimensions of context that matter in learning. Our use of AI needs to follow this trajectory toward context to support educational applications.

To achieve human-centric vision, listening session attendees argued that teams will need time and freedom to explore how best to manage the tension between the pace of technological

advancement and the need for broader contextual insights—for trust and for safety. They will need time and freedom to pioneer new processes that better involve teachers and students as co-designers, with attention to balancing power dynamics. And they will need to shift attention from older ways of framing priorities (such as achievement gaps) to new ways of prioritizing digital equity. We call on R&D funders to focus resources on the long tail of learner variability, the need for AI-enabled systems that better incorporate context, and time required to get contextual considerations right. We call upon researchers and developers to prioritize challenges of context, trust, and safety in their work to advance AI.

Recommendation #7: Develop Education-Specific Guidelines and Guardrails

Our final recommendation is central to policymakers. A feature of the American educational system is the emphasis on local decision making. With technology growing in complexity at such a rapid pace, it is becoming difficult for local leaders to make informed decisions about the deployment of artificial intelligence. As we have discussed, the issues are not only data privacy and security but extend to new topics such as bias, transparency, and accountability. It will be harder to evaluate promising edtech platforms that rely on AI systems against this evolving, complex set of criteria.

Regulations related to key student and family data privacy laws like the Family Educational Rights & Privacy Act (FERPA), the Children's Internet Privacy Act (CIPA), and the Children's Online Privacy Protection Act (COPPA) warrant review and further consideration in light of new and emerging technologies in schools. Laws such as the Individuals with Disabilities Education Act (IDEA) may likewise be considered as new situations arise in the use of AI-enabled learning technologies. As discussed throughout this document, the *Blueprint for an AI Bill of Rights* is an important framework throughout this work.

The Department encourages parallel work by constituents in all levels of the educational system. In addition to the key federal laws cited immediately above, many states have also passed privacy laws that govern the use of educational technology and edtech platforms in classrooms. Further constituents can expect general frameworks for responsible AI in parallel sectors like health, safety, and consumer products to be informative but not sufficient for education's specific needs. Leaders at every level need awareness of how this work reaches beyond implications for privacy and security (e.g., to include awareness of potential bias and unfairness), and they need preparation to effectively confront the next level of issues.

Next Steps

We are heartened to see intensifying discussions throughout the educational ecosystem about the role of AI. We see progress that we can build upon occurring, as constituents discuss these three types of questions: What are the most significant opportunities and risks? How can we achieve trustworthy educational AI? How can we understand the models at the heart of applications of AI and ensure they have the qualities that align to educational aspirations?

The Department developed this report with awareness of contributions arising from many types of organizations and collectives. Internationally, we recognize parallel efforts to consider AI in the European Union, at the United Nations, and indeed throughout the world. We are aware of progress being led by organizations such as UNESCO, the EdSAFE AI Alliance, and research

organizations in many countries. We plan to continue cross-agency work, for example, by continuing to coordinate with the Office of Science and Technology Policy and other Federal agencies as agencies implement next steps guided by the *Blueprint for an AI Bill of Rights*. We see a broad and fertile context for necessary next steps:

- Working within this context and with others, the Department will consider specific policies and regulations so that educators can realize the opportunities of AI in edtech while minimizing risks. For example, the Department is developing a set of AI usage scenarios to strengthen the process of evaluating and enhancing policies and regulations. The principles and practices in the *Blueprint for an AI Bill of Rights* will be used to ensure the scenarios mitigate important risks and harms.

- Working with constituents (including education leaders; teachers, faculty, support staff, and other educators; researchers; policymakers; funders; technology developers; community members and organizations; and above all, learners and their families/caregivers), we will develop additional resources and events to increase understanding of AI and to involve those who will be most affected by these new technologies.

- Working across sectors, such as education, innovation, research, and policy, we will revise and update the NETP to guide all constituents toward safe, equitable, and effective AI in education in the United States, in alignment with our overall educational priorities.

Common Acronyms and Abbreviations

- AES: Automated Essay Scoring
- AI: Artificial Intelligence
- CIPA: Children's Internet Protection Act
- COPPA: Children's Online Privacy Protection Act
- Edtech: Educational Technology
- ESEA: Elementary and Secondary Education Act
- ESSA: Every Student Succeeds Act
- FERPA: Family Educational Rights and Privacy Act
- IA: Intelligence Augmentation
- IDEA: Individuals with Disabilities Education Act
- IEP: Individualized Education Program
- ITS: Intelligent Tutoring Systems
- NETP: National Education Technology Plan
- R&D: Research & Development

Acknowledgements

Project Team

Artificial Intelligence and the Future of Teaching and Learning was developed under the leadership and guidance of **Roberto J. Rodríguez,** Assistant Secretary for the Office of Planning, Evaluation, and Policy Development, **Kristina Ishmael**, Deputy Director of the Office of Educational Technology, and **Bernadette Adams**, Senior Policy Advisor for the Office of Educational Technology at the U.S. Department of Education.

Support for the creation of this document was provided by Digital Promise, led by **Jeremy Roschelle** with **Carly Chillmon, Judi Fusco, Gabrielle Lue, Eric Nentrup, My Nguyen, Pati Ruiz, and Zohal Shah**. Special thanks to Center for Integrative Research in Computing and Learning Sciences postdocs **Michael Chang** and **Aditi Mallavarapu**.

Listening Session Panelists and Hosts

Hal Abelson

Ryan Baker

Nancye Blair Black

Marcelo Aaron Bonilla Worsley

Michael Chang

Carly Chillmon

Sherice Clarke

Tammy Clegg

Sidney d'Mello

Judi Fusco

Dragan Gasevic

Kip Glazer

Janice Gobert

Sarah Hampton

Kristina Ishmael

Jim Larimore

Nicol Turner Lee

Sherry Loftin

Gabrielle Lue

Aditi Mallavarapu

Ole Molvig

Peter Norvig

Thomas Philip

Vidula Plante

Jeremy Roschelle

Pati Ruiz

Alina Von Davier

Erin Walker

Diego Zapata

We also thank 1,075 people who registered for Listening Sessions and 700 who attended.

References

Akgun, S., Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI Ethics, 2,* 431–440. https://doi.org/10.1007/s43681-021-00096-7

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction based on adaptive learning technologies. In Mayer, R.E. & Alexander, P.A., *Handbook of research on learning and instruction,* 522-560. ISBN: 113883176X

Baker, R.S., Esbenshade, L., Vitale, J., & Karumbaiah, S. (2022). Using demographic data as predictor variables: A questionable choice. https://doi.org/10.35542/osf.io/y4wvj

Black, P. & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 92(1),* 81-90. https://kappanonline.org/inside-the-black-box-raising-standards-through-classroom-assessment/

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21(1),* 5-31. https://doi.org/10.1007/s11092-008-9068-5

Boden, M.A. (2018). *Artificial intelligence: A very short introduction.* Oxford. ISBN: 978-0199602919

Bryant, J., Heitz,C., Sanghvi, S., & Wagle, D. (2020, January 14). *How artificial intelligence will impact K-12 teachers.* McKinsey. https://www.mckinsey.com/industries/education/our-insights/how-artificial-intelligence-will-impact-k-12-teachers

Celik, I., Dindar, M., Muukkonen, H. & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends, 66,* 616–630. https://doi.org/10.1007/s11528-022-00715-y

Center for Integrative Research in Computing and Learning Sciences (CIRCLS). (2022, Feb.). *From Broadening to empowering: Reflecting on the CIRCLS'21 Convening.* https://circls.org/circls21report

Chen, C., Park, H.W. & Breazeal, C. (2020). Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Computers & Education, 150,* https://doi.org/10.1016/j.compedu.2020.103836

Chen, G., Clarke, S., & Resnick, L.B. (2015). Classroom Discourse Analyzer (CDA): A discourse analytic tool for teachers. *Technology, Instruction, Cognition and Learning, 10(2),* 85-105

Dieterle, E., Dede, C. & Walker, M. (2022). The cyclical ethical effects of using artificial intelligence in education. *AI & Society.* https://link.springer.com/article/10.1007/s00146-022-01497-w

Doewes, A. & Pechenizkiy, M. (2021). On the limitations of human-computer agreement in automated essay scoring. In *Proceedings of the 14th International Conference on Educational Data Mining* (EDM21). https://educationaldatamining.org/EDM2021/virtual/static/pdf/EDM21_paper_243.pdf

Englebart, D.C. (October 1962). *Augmenting human intellect: A conceptual framework*. SRI Summary Report AFOSR-3223. https://www.dougengelbart.org/pubs/augment-3906.html

Ersozlu, Z., Ledger, S., Ersozlu, A., Mayne, F., & Wildy, H. (2021). Mixed-reality learning environments in teacher education: An analysis of TeachLivETM Research. *SAGE Open, 11(3).* https://doi.org/10.1177/21582440211032155.

European Commission, Directorate-General for Education, Youth, Sport and Culture. (2022). *Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators.* Publications Office of the European Union. https://data.europa.eu/doi/10.2766/153756

Forsyth, S., Dalton, B., Foster, E.H., Walsh, B., Smilack, J., & Yeh, T. (2021, May). Imagine a more ethical AI: Using stories to develop teens' awareness and understanding of artificial intelligence and its societal impacts. In *2021 Conference on Research in Equitable and Sustained Participation in Engineering, Computing, and Technology* (RESPECT). IEEE. https://doi.org/10.1109/RESPECT51740.2021.9620549

Friedman, L., Blair Black, N., Walker, E., & Roschelle, J. (November 8, 2021) Safe AI in education needs you. *Association of Computing Machinery BLOG@ACM*, https://cacm.acm.org/blogs/blog-cacm/256657-safe-ai-in-education-needs-you/fulltext

Gardner, J., O'Leary, M. & Yuan, L. (2021). Artificial intelligence in educational assessment: "Breakthrough? Or buncombe and ballyhoo?" *Journal of Computer Assisted Learning, 37(5),* 1207–1216. https://doi.org/10.1111/jcal.12577

Gartner (n.d.) *Gartner glossary: Augmented intelligence.* Gartner. https://www.gartner.com/en/information-technology/glossary/augmented-intelligence

Gay, G. (2018). *Culturally responsive teaching: Theory, research, and practice.* Teachers College Press. ISBN: 978-0807758762

Godwin-Jones, R. (2021). Big data and language learning: Opportunities and challenges. *Language Learning & Technology, 25(1),* 4–19. http://hdl.handle.net/10125/44747

Hammerness, K., Darling-Hammond, L., & Bransford, J. (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do.* Jossey-Bass. ISBN: 0787996343

Holmes, W. & Porayska-Pomsta, K. (Eds.) (2022). *The ethics of artificial intelligence in education. Routledge.* ISBN 978-0367349721

Holstein, K., McLaren, B.M., & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher–AI complementarity. *Journal of Learning Analytics, 6(2).* https://doi.org/10.18608/jla.2019.62.3

IEEE-USA Board of Directors. (February 10, 2017). *Artificial intelligence research, development and regulation.* IEEE http://globalpolicy.ieee.org/wp-content/uploads/2017/10/IEEE17003.pdf

Jensen, E., Dale, M., Donnelly, P.J., Stone, C., Kelly, S., Godley, A. & D'Mello, S.K. (2020). Toward automated feedback on teacher discourse to enhance teacher learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20). https://doi.org/10.1145/3313831.3376418

Kai, S., Almeda, M.V., Baker, R. S., Heffernan, C., & Heffernan, N. (2018). Decision tree modeling of wheel-spinning and productive persistence in skill builders. *Journal of Educational Data Mining, 10(1),* 36–71. https://doi.org/10.5281/zenodo.3344810

Kaplan, R.M., & Saccuzzo, D.P. (2017). *Psychological testing: Principles, applications, and issues.* Cengage Learning.

Ke, Z., & Ng, V. (2019). Automated essay scoring: A survey of the state of the art. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 6300–6308. https://doi.org/10.24963/ijcai.2019/879

Khosravi, H., Shum, S.B., Chen, G, Conati, C., Tsai,Y-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., Gašević, D. (2022). Explainable artificial intelligence in education. Computers and Education: *Artificial Intelligence, 3.* https://doi.org/10.1016/j.caeai.2022.100074

Kulik, J.A., & Fletcher, J.D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research, 86(1),* 42–78

Ma, W., Adescope, O.O, Nesbit, J.C. & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology, 106(4),* 901–918. http://dx.doi.org/10.1037/a0037123

Maslej, N., Fattorini, L., Brynjolfsson E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J.C., Parli, V., Shoham, Y., Wald, R., Clark, J. and Perrault, R., (2023). *The AI index 2023 annual report.* Stanford University: AI Index Steering Committee, Institute for Human-Centered AI.

Merrill, S. (2020). *In schools, are we measuring what matters?* Edutopia. https://www.edutopia.org/article/schools-are-we-measuring-what-matters

Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education, 00, 1–14.* https://doi.org/10.1111/ejed.12527

Mostow, J., Aist, G., Burkhead, P., Corbett, A., Cuneo, A., Eitelman, S., Huang, C., Junker, B., Sklar, M.B., & Tobin, B. (2003). Evaluation of an automated reading tutor that listens: Comparison to human tutoring and classroom instruction. *Journal of Educational Computing Research, 29(1),* 61–117. https://doi.org/10.2190/06AX-QW99-EQ5G-RDCF

Mousavinasab, E., Zarifsanaiey, N., R. Niakan Kalhori, S., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments, 29(1),* 142–163. https://psycnet.apa.org/doi/10.1080/10494820.2018.1558257

National Academies of Sciences, Engineering, and Medicine. 2018. *How people learn II: Learners, contexts, and cultures.* The National Academies Press. https://doi.org/10.17226/24783

National Research Council. 2000. *How people learn: Brain, mind, experience, and school.* The National Academies Press. https://doi.org/10.17226/9853

Nentrup, E. (2022). *How Policymakers Can Support Educators and Technology Vendors Towards SAFE AI.* EdSAFE AI Alliance. https://www.edsafeai.org/post/how-policymakers-can-support-aied

Page, E.B. (1966). The imminence of grading essays by computer. *Phi Delta Kappan, 47(5),* 238–243

Paris, D., & Alim, H.S. (Eds.). (2017). *Culturally sustaining pedagogies: Teaching and learning for justice in a changing world.* Teachers College Press. ISBN: 978-0807758342

Plass, J.L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education, 52(3),* 275–300. https://doi.org/10.1080/15391523.2020.1719943

Regona, Massimo & Yigitcanlar, Tan & Xia, Bo & Li, R.Y.M. (2022). Opportunities and adoption challenges of AI in the construction industry: A PRISMA review. *Journal of Open Innovation Technology Market and Complexity, 8(45).* https://doi.org/10.3390/joitmc8010045

Reynolds, C.R., & Suzuki, L.A. (2012). Bias in psychological assessment: An empirical review and recommendations. *Handbook of Psychology, Second Edition*. https://doi.org/10.1002/9781118133880.hop210004

Ritter, S., Anderson, J.R., Koedinger, K.R. & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review, 14,* 249–255/ https://doi.org/10.3758/BF03194060

Roll, I., Aleven, V., McLaren, B.M., Koedinger, K.R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system, *Learning and Instruction, 21(2),* 267–280. https://doi.org/10.1016/j.learninstruc.2010.07.004.

Roschelle, J., Dimitriadis, Y. & Hoppe, U. (2013). Classroom orchestration: Synthesis. *Computers & Education, 69,* 512-526. https://doi.org/10.1016/j.compedu.2013.04.010

Roschelle, J., Feng, M., Murphy, R. & Mason, C.A. (2016). Online mathematics homework increases student achievement. *AERA Open, 2(4),* 1-12. DOI: 10.1177/2332858416673968

Roschelle, J., Penuel, W., & Shechtman, N. (2006). Co-design of innovations with teachers: definition and dynamics. In *Proceedings of the 7th International Conference on Learning Sciences,* Bloomington, IN. https://doi.dx.org/10.22318/icls2006.606

Rose, D. (2000). Universal design for learning. *Journal of Special Education Technology, 15(4),* 47-51. https://doi.org/10.1177/016264340001500407

Ruiz, P. & Fusco, J. (2022). *Teachers partnering with artificial intelligence: Augmentation and automation.* Digital Promise. https://digitalpromise.org/2022/07/06/teachers-partnering-with-artificial-intelligence-augmentation-and-automation/

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control.* Viking. ISBN 978-0-525-55861-3.

Shao, Q., Sniffen, A., Blanchet, J., Hillis, M.E., Shi, X., Haris, T.K., & Balkcom, D. (2020). Teaching american sign language in mixed reality. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(4),* 1-27. https://doi.org/10.1145/3432211

Sharples, M. & Pérez y Pérez, R. (2022). *Story machines: How computers have become creative writers.* Routledge. ISBN 9780367751951

Shemshack, A., Spector, J.M. (2020) A systematic literature review of personalized learning terms. *Smart Learning Environments, 7(33).* https://doi.org/10.1186/s40561-020-00140-9

Shute, V J. (2008). Focus on formative feedback. *Review of Educational Research, 78(1),* 153–189. https://doi.org/10.3102/0034654307313795

Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and learning of qualitative physics in Newton's Playground. *The Journal of Educational Research, 106(6),* 423-430. https://doi.org/10.1080/00220671.2013.832970

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J.M., Milligan, S., Selwyn, B. & Gašević,D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence, 3.* https://doi.org/10.1016/j.caeai.2022.100075

The White House (February 17, 2023). *Executive order on further advancing racial equity and support for underserved communities through the federal government.* https://www.whitehouse.gov/briefing-room/presidential-actions/2023/02/16/executive-order-on-further-advancing-racial-equity

The White House (September 8, 2022). *Readout of White House listening session on tech platform accountability*. https://www.whitehouse.gov/briefing-room/statements-releases/2022/09/08/readout-of-white-house-listening-session-on-tech-platform-accountability/

U.S. Department of Education, Office of Educational Technology (2022). *Advancing digital equity for all: Community-based recommendations for developing effective digital equity plans to close the digital divide and enable technology-empowered learning.* US Department of Education.

U.S. Department of Education, Office of Educational Technology. (2010). *Transforming American Education: Learning Powered by Technology.* U.S. Department of Education. p. 78

Van Lehn, K. (2011) The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46(4),* 197-221. https://doi.org/10.1080/00461520.2011.611369

Wagner, A.R., Borenstein, J. & Howard, A. (September 2018). Overtrust in the robotics age. *Communications of the ACM, 61(9),* 22-24. https://doi.org/10.1145/3241365

Walker, E., Rummel, N. & Koedinger, K.R. (2015). Adaptive intelligent support to improve peer tutoring in algebra. *International Journal of Artificial Intelligence in Education, 24,* 33–61 https://doi.org/10.1007/s40593-013-0001-9

Walton Family Foundation (March 1, 2023). *Teachers and students embrace ChatGPT for education.* https://www.waltonfamilyfoundation.org/learning/teachers-and-students-embrace-chatgpt-for-education

Webb, N.M., & Farivar, S. (1994). Promoting helping behavior in cooperative small groups in middle school mathematics. *American Educational Research Journal, 31(2),* 369–395. https://doi.org/10.3102/00028312031002369

White House Office of Science and Technology Policy (October 2022), *Blueprint for an AI bill of rights: Making automated systems work for the American people.* The White House Office of Science and Technology Policy. https://www.whitehouse.gov/ostp/ai-bill-of-rights/

Wiggins, G. (2015). *Seven keys to effective feedback.* ACSD. https://www.ascd.org/el/articles/seven-keys-to-effective-feedback

Winne, P.H. (2021). Open learner models working in symbiosis with self-regulating learners: A research agenda. *International Journal of Artificial Intelligence in Education, 31(3),* 446-459. https://doi.org/10.1007/s40593-020-00212-4

Zacamy, J. & Roschelle, J. (2022). *Navigating the tensions: How could equity-relevant research also be agile, open, and scalable?* Digital Promise. http://hdl.handle.net/20.500.12265/159

Zhai, X., He, P., Krajcik, J. (2022). Applying machine learning to automatically assess scientific models. *Journal of Research in Science Teaching.* https://doi.org/10.1002/tea.21773

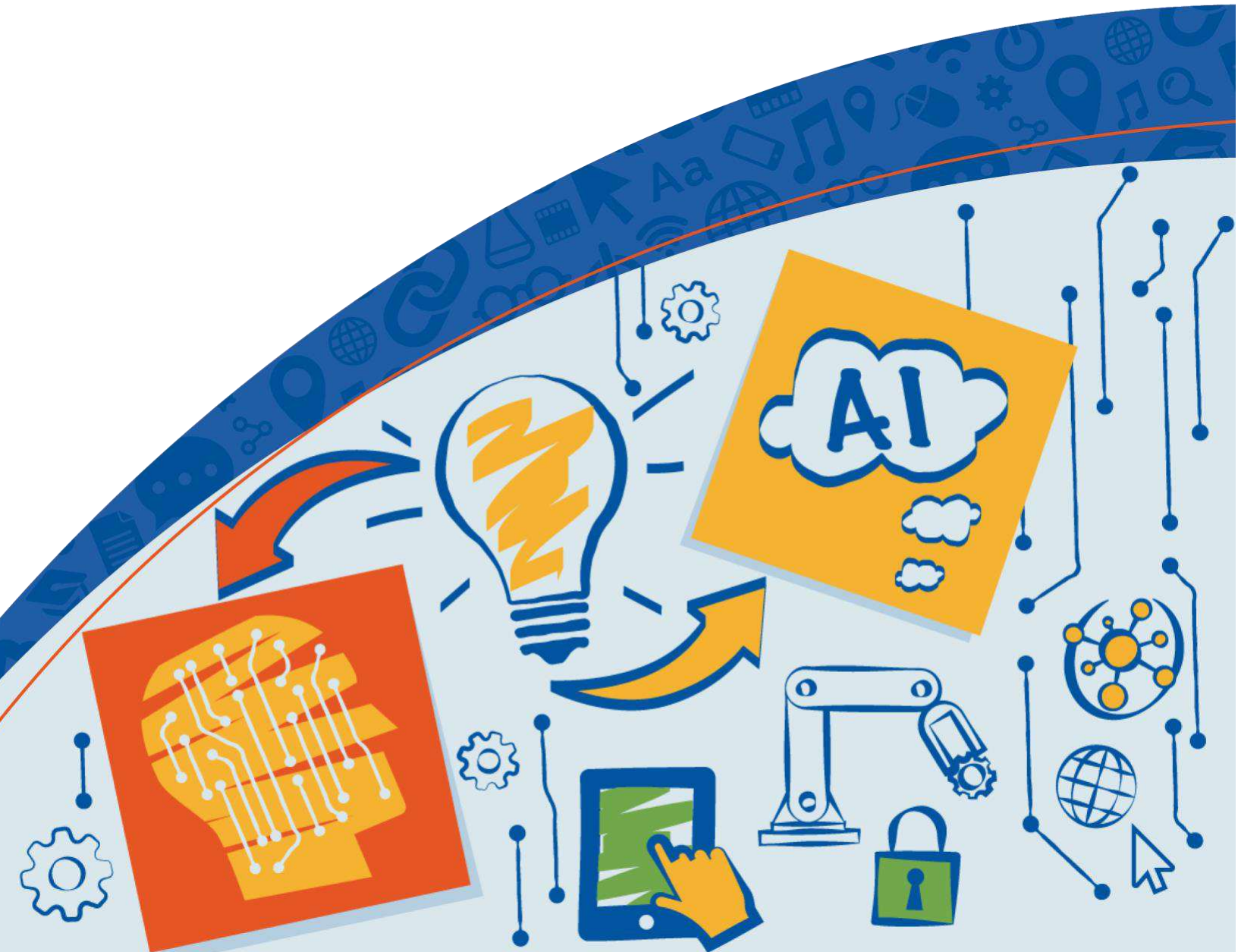
Zhang, H., Lee, I., Ali, S., DiPaola, D., Cheng, Y., & Breazeal, C. (2022). Integrating ethics and career futures with technical learning to promote AI literacy for middle school students: An exploratory study. *International Journal of Artificial Intelligence in Education*, 1–35. https://doi.org/10.1007/s40593-022-00293-3

OFFICE OF
Educational Technology

Designing for Education with Artificial Intelligence:

An Essential Guide for Developers

July 2024



Designing for Education with Artificial Intelligence: An Essential Guide for Developers

Miguel A. Cardona, Ed.D.

Secretary, U.S. Department of Education

Roberto J. Rodriguez

Assistant Secretary, Office of Planning, Evaluation and Policy Development

July 2024

Examples Are Not Endorsements

This document contains examples and resource materials that are provided for the user's convenience. The inclusion of any material is not intended to reflect its importance, nor is it intended to endorse any views expressed or products or services offered. These materials may contain the views and recommendations of various subject matter experts as well as contact addresses, websites, and hypertext links to information created and maintained by other public and private organizations. The opinions expressed in any of these materials do not necessarily reflect the positions or policies of the U.S. Department of Education. The U.S. Department of Education does not control or guarantee the accuracy, relevance, timeliness, or completeness of any information from other sources that are included in these materials. Other than statutory and regulatory requirements included in the document, the contents of this guide do not have the force and effect of law and are not meant to bind the public.

Contracts and Procurement

This document is not intended to provide legal advice or approval of any potential federal contractor's business decision or strategy in relation to any current or future federal procurement and/or contract. Further, this document is not an invitation for bid, request for proposal, or other solicitation.

Licensing and Availability

Notice to Persons with Limited English Proficiency

If you have difficulty understanding English, you may request language assistance services for U.S. Department of Education information that is available to the public. These language assistance services are available free of charge. If you need more information about interpretation or translation services, please call 1-800-USA-LEARN (1-800-872-5327) (TTY: 1-800-437-0833); email us at Ed.Language.Assistance@ed.gov; or write to U.S. Department of Education, Information Resource Center, LBJ Education Building, 400 Maryland Ave. SW, Washington, DC 20202.

How to Cite

Acknowledgements

Project Team

Designing for Education with Artificial Intelligence: An Essential Guide for Developers was developed under the leadership and guidance of **Roberto J. Rodríguez**, Assistant Secretary for the Office of Planning, Evaluation and Policy Development, **Anil Hurkadli**, Interim Deputy Director for the Office of Educational Technology, **Bernadette Adams**, Senior Policy Advisor for the Office of Educational Technology, and **Kevin Johnstun**, Education Program Specialist for the Office of Educational Technology.

This work was developed with support from Digital Promise under contract (91990019C0076) and led by **Jeremy Roschelle** with **Anthony Baker**, **Pati Ruiz**, **Eric Nentrup, Gabrielle Lue** and **Sarah Martin**.

Contributing Members of the Developer Community

- Kristen DiCerbo, Khan Academy
- Teddy Hartman, GoGuardian
- Neil Heffernan, Worcester Polytechnic Institute
- Karl Rectanus, EDSAFE AI Alliance
- Steve Ritter, Carnegie Learning
- Sharad Sundararajan, Merlyn Mind
- Alyssa Van Camp, TeachFx
- Julia Winter, Alchemie Solutions, Inc.

We also thank the many developers, industry associations, and nonprofit organizations that attended our listening sessions and contributed their ideas for translating the Department's recommendations for Artificial Intelligence in education into practical guidelines.

Table of Contents

Designing for Education with Artificial Intelligence: An Essential Guide for Developers

Introduction

Today and in the future, a growing array of Artificial Intelligence (AI) models and capabilities will be incorporated into the products that specifically serve educational settings. The U.S. Department of Education (Department) is committed to encouraging innovative advances in educational technology (edtech) to improve teaching and learning across the nation's education systems and to supporting developers as they create products and services using AI for the educational market.

Building on the Department's prior report, *Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations* (2023 AI Report), this guide seeks to inform product leads and their teams of innovators, designers, developers, customer-facing staff, and legal teams as they work toward safety, security, and trust while creating AI products and services for use in education. This landscape is broader than those building large language models (LLMs) or deploying chatbots; it includes all the ways existing and emerging AI capabilities can be used to further shared educational goals.

Our insights here are intended to support people who are managing teams in the design and development of products that leverage AI to improve teaching and learning. We have attempted to address topics that will be relevant across the continuum of edtech developers, which includes established firms and newcomers, as well as developers across research, nonprofit, and for-profit organizations. We address not only developers of products for formal education settings—including elementary and secondary schools, colleges, and universities—but also for educational uses at home, community, and other informal settings.

To this end, each section of this document is built around a core recommendation and includes a set of discussion questions that leaders in organizations can use to foster conversation, next steps to promote robust development processes, and resources that can provide additional support. Please note that the Department and other federal agencies are actively considering next steps to promote the safe and responsible use of AI. Thus, this document suggests "questions to ask" and "directions to pursue" to developers that are deliberately open-ended.

This guide provides non-regulatory, education-specific guidance that is aligned with federal guidelines and guardrails. This guide's coverage of existing federal guidelines and guardrails is not comprehensive or exhaustive. It is not intended to and does not enable a developer to establish its compliance with regulations. Also, it is not intended to and does not introduce any new requirements. Where examples are given, including links to non-U.S. Government websites, they are intended to be illustrative and not to restrict the application of this guide to additional forms of AI as they become available for use in education. We are providing these external links because they contain additional information relevant to the topic(s) discussed in this document or that

otherwise may be useful to the reader. We cannot attest to the accuracy of information provided on the cited third-party websites or any other linked third-party site. We are providing these links for reference only; linking to external resources does not constitute an endorsement by the Department. Developers can use this guide to increase their understanding of essential federal guidelines and guardrails to guide their work as they create AI applications for educational settings.

Responding to the October 2023 Executive Order

This guide is responsive to President Joe Biden's October 30, 2023, Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (Executive Order on AI), which states the following:

> *To help ensure the responsible development and deployment of AI in the education sector, the Secretary of Education shall, within 365 days of the date of this order, develop resources, policies, and guidance regarding AI. These resources shall address safe, responsible, and nondiscriminatory uses of AI in education, including the impact AI systems have on vulnerable and underserved communities, and shall be developed in consultation with stakeholders as appropriate.*

This guide is informed by an extensive series of public listening sessions with students, parents, and educators along with developers, industry associations, and nonprofit organizations. This included a cross section of developers representing a variety of company sizes, funding models, and organization types (for-profit/nonprofit). Session participants shared their current approaches to safety and security, the risks they and their users face, suggestions on supports and resources, and thoughts about opportunities to build trust in the future. Several additional listening sessions occurred with a smaller set of developers (listed in the contributing members section above) drawn from those who participated initially. Where this guide refers to listening sessions, it includes all these opportunities to hear from constituents.

This guide draws on a growing series of federal publications on AI, which includes these examples:

- **Office of Science and Technology Policy Blueprint for an AI Bill of Rights**: The 2022 White House white paper that broadly shapes a strategy informing and protecting citizens from AI and related technologies
- **National Institute of Standards and Technology (NIST) AI Risk Management Framework**: A seven-step framework to protect developer and end user interests alike, particularly relevant to edtech companies using AI components and forthcoming emerging technologies
- **FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Eight Additional Artificial Intelligence Companies to Manage the Risks Posed by AI | The White House**: The White House's engagement with technology companies highlights shared responsibility among the federal government, developer organizations, and other constituents.

Defining "Artificial Intelligence" and "EdTech" Broadly

The Department takes a broad view of the terms "AI" and "edtech." This document's guidance applies broadly across the many types of AI that developers may integrate and the many ways their products may be used in educational settings.

Box A: Artificial Intelligence as defined in the Executive Order on AI.

The term "artificial intelligence" means a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to perform the following:

- Perceive real and virtual environments
- Abstract such perceptions into models through analysis in an automated manner
- Use model inference to formulate options for information or action

As a starting point for defining "AI," we note the statutory definition of AI (see Box A above), which also appears in the Executive Order on AI. As we noted in the Department's 2023 AI Report, AI is an umbrella term for many subfields of research and innovation. Following the 2023 AI Index Report, from the National Artificial Intelligence Research Resource Task Force, we continue to observe the extremely rapid evolution of AI capabilities in many domains, such as speech, vision, robotics, and text.

Box B: Defining Educational Technology (edtech)

As defined in the Department's **2023 AI Report**, edtech includes the following:

A. Technologies specifically designed for educational use

B. General technologies that are widely used within educational settings

Likewise, we define "edtech" broadly (Box B). In an independent market research firm's report, the global edtech market for both K-12 and higher education is valued at $123 billion as of 2022. This market includes startups, small businesses, nonprofit organizations, and larger corporations—companies with a mission specific to education and others with products that are used in education and other sectors. There is a broad spectrum of products ranging from infrastructure to student information systems to learning management systems to specific end-user applications and more.

Whereas a market is defined by buyers and sellers, we define the edtech "ecosystem" broadly to include the many different people and organizations working together to design and refine new products and services. Within this ecosystem, discussions of the Executive Order concepts of "safety, security, and trust" will be infused with concepts more specific to education (e.g., evidence, fairness, data privacy) as elaborated below. Ecosystem participants include

educational procurement departments, additional educational decision makers, educators, parents and guardians, students, nonprofits, postsecondary education institutions, and broader community members. The ecosystem includes people who directly create or use AI products and people involved in educational systems who are affected by AI products. Shared responsibility for building trust goes well beyond buying and selling; it thrives on open and engaged communication across the ecosystem.

Key Message: Shared Responsibility for Building Trust

Through listening sessions, the Department learned that one key recommendation in the 2023 AI Report, "Prioritize Strengthening Trust," resonated with developers as a call to action. Developers recognize the fundamental importance of "trust" in the edtech ecosystem in which they participate. Trust improves the co-design process between developers and educators so that together they can create and scale innovative products. Consequently, developers can benefit greatly from understanding how they can work with others in the ecosystem to strengthen trust. Specific key messages follow:

1. Trust is a shared responsibility.

The President's Executive Order on AI makes clear beginning with its title that the federal government has a responsibility to promote "safe, secure, and trustworthy development" and articulates a stance of shared responsibility (see quote below). Consequently, developers will find information within this document on where to locate federal laws and other federal resources that are directly applicable to their work in education. Because the technology is evolving rapidly, developers may find it valuable to go beyond attending to and complying with today's federal guidelines and guardrails to earn trust. One important example is the Software and Information Industry Association's Principles for the Future of AI in Education, which articulates seven principles (e.g., their evaluation principle calls for continually assessing the impact of AI). TeachAI has also developed Principles for AI in Education as part of its AI Guidance for Schools Toolkit to guide the effective development and application of AI in teaching and learning. As a third example, many developers have participated in the EDSAFE AI Alliance, which has produced a common framework, the SAFE Benchmarks.

> *"Harnessing AI for good and realizing its myriad benefits requires mitigating its substantial risks. This endeavor demands a society-wide effort that includes government, the private sector, academia, and civil society."*
>
> *— Executive Order on AI*

Other members of the edtech ecosystem also are accepting responsibility and developing conditions for trust. Educational leaders at all levels, including state, district, and building-level leaders, are writing their own guidance (Box C). Developers can look to these resources to understand the steps that educators are taking to build understanding and capacity; to strengthen procurement processes; to protect privacy, security, and fairness; and to manage other forms of risk. Further, in a forthcoming Toolkit for Educational Leaders developed

pursuant to the Executive Order on AI, the Department will be shaping the shared responsibility conversation among educators. Many nonprofits are also developing helpful toolkits and resources (see Box D). As key participants in the edtech ecosystem, developers are encouraged to interact responsibly with the ecosystem to develop trust.

Box C: State guidance resources about AI in education, as of June 2024

Arizona	Kentucky	Oregon
California	Mississippi	Utah
Connecticut	North Carolina	Virginia
Hawaii	Ohio	Washington State
Indiana	Oklahoma	West Virginia

Box D: Partial listing of nonprofits offering guidance resources about AI in education

- The Consortium for School Networking (CoSN) Gen AI Maturity Tool
- EDSAFE AI Alliance SAFE Benchmarks Framework
- International Society for Technology and Education 's Artificial Intelligence in Education Resource List
- Software & Information Industry Association's Principles for AI in Education
- TeachAI's AI Guidance for Schools Toolkit
- All4Ed's Future Ready Schools Emerging Practices Guide

2. Trust requires actively managing AI risks so that we can seize its benefits.

Through its conversations with developers, the Department observed an important shift in how developers are engaging with others around their work. Whereas it has been common to present "solutions"—how technologies can improve teaching, learning, and other educational processes—developers are now also openly discussing how they are managing risks. Some developers have publicly shared details on the process they went through to identify, prioritize, and manage risks. As developers openly discuss risk management, the Department suggests attending to two kinds of processes: (a) technical development processes that result in *trustworthy* systems and (b) engagement strategies that *build trusting relationships* among developers and other ecosystem members.

Thus, the Department understands the importance of discussing both opportunities and risks in a responsible manner. Box E lists examples of salient categories of risk.[1,2,3] Both risks and opportunities will be discussed in more depth later in this guide.

Box E: Types of risks of AI, ordered alphabetically, not by priority

- AI "Race-to-Release" Risks
- Bias and Fairness Risks
- Data Privacy and Security Risks
- Harmful Content Risks
- Ineffective System Risks
- Malicious Use Risks
- Misinformation Management Risks (including "hallucinations")
- Transparency and Explainability Risks
- Underprepared User Risks

In considering risks, it is important to note that AI is evolving rapidly. For example, just as educators are gaining familiarity with text-oriented chatbots, industry is advancing and releasing multimodal capabilities that add new layers to the potential risks. Hence, this guide outlines risks broadly and asks developers to adopt risk mitigation processes that address both the risks that are foreseeable today and those that will newly emerge.

Risks are not only intrinsic to the technology; risks also emerge at the interface of technology and human activity. As people use AI, both foreseeable and unforeseen risks will arise. In its 2023 AI Report, the Department recommended "humans in the loop" and yet, asking an educator to review every use of AI or every AI-based output is neither practical nor fair. Developers share responsibility to be "in the loop" to review uses and outputs of AI, both during the development process and as a product is used in the field. Building on Box E, we illustrate challenging scenarios where both developers and educators will likely need to attend to emergent risks, with a division of responsibility that is yet to be determined:

[1] *Potential Risks of Artificial Intelligence Integration into School Education: A Systematic Review.*

[2] *Getting To Know—and Manage—Your Biggest AI Risks.*

[3] *The Promises and Perils of Generative AI in Education: TFA's Evolving Perspective.*

- As teachers generate personalized lesson plans with AI services, who will review and revise the outputs to eliminate false information generated by large language models and confirm the content is accurate and aligned to educational objectives?
- As curriculum coordinators engage AI to support their work in curating instructional resources and formative assessments for use in their schools, who will weigh evidence for the efficacy of the resources and the validity of the assessments? Who will verify that resources address the needs of underserved and vulnerable populations?
- As guidance counselors use AI-assisted tools to recommend college and career pathways, who will detect and counter unfairness in the recommendations due to biases in historical data sets that were used to develop the AI model and which could harm vulnerable populations?
- As educators use AI to simplify their work of writing emails or other correspondence about their students' work, who is responsible for safeguards against disclosing a student's private information to unintended recipients, including the developer of the AI model?
- As administrators and school leaders procure early warning systems to identify students who may be "at risk," who has sufficient knowledge and time to evaluate whether the AI developer adhered to scientific, legal, and privacy standards necessary to safeguard students' civil rights?
- As educators deploy anti-plagiarism detectors to identify a student's inappropriate use of edtech, who has responsibility for recognizing weaknesses and biases in AI-based detectors that could lead to disciplining students unfairly or unequally? Who ensures that underserved and vulnerable populations are not unfairly targeted?

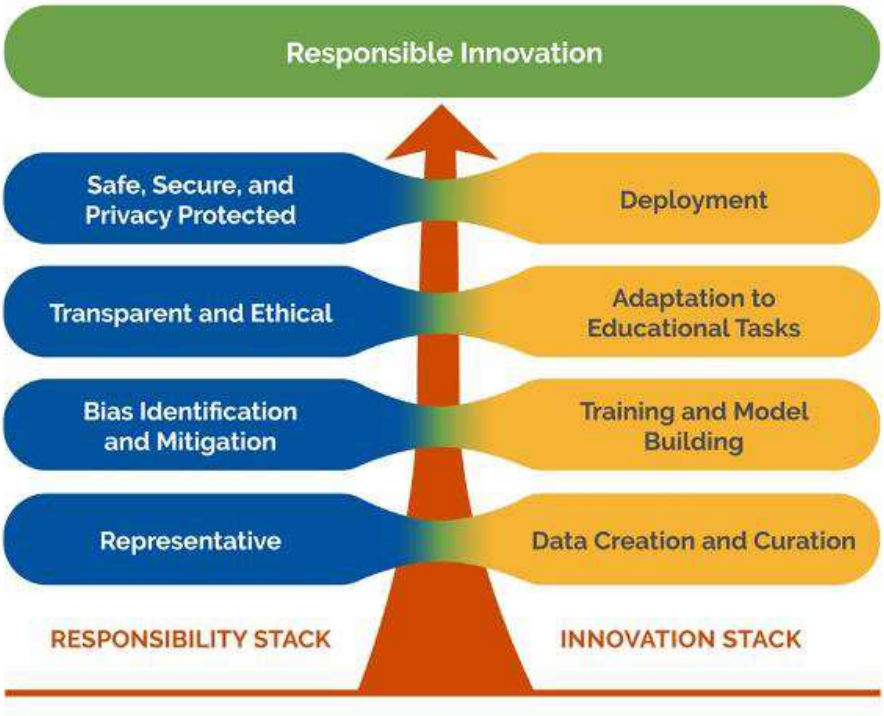
In each challenging case above, the Department respects the role of educators in overseeing educational decisions. However, AI developers should assume significant responsibility because it is unreasonable to ask educators to be primary reviewers of the data and methods used to develop AI models and related software. Building on an important emphasis of the Executive Order on AI, the Department calls on developers to pay special attention to identifying and mitigating potential harms to underserved and vulnerable populations from AI.

3. Coordinating Innovation and Responsibility Throughout Development: A "Dual Stack"

Developers use the visual metaphor of a stack to describe how products are built from layered or connected components. A development stack is a powerful way to coordinate the work of many innovators in a complex organization; it coordinates how the overall product or service will be produced and delivered to market. The Department strongly encourages development organizations to now define a parallel strength via a coordinated "responsibility" stack. This stack would establish how people in a complex edtech development organization work together to earn the trust of educational users of their products. See Figure 1.

Figure 1: Developers should integrate a Responsibility Stack with their Innovation Stack.

The specific elements of each stack are illustrative, not comprehensive. Each developer organization should elaborate the stacks to fit the specifics of their efforts.



Indeed, whereas developers may have previously emphasized one particular risk management office or role in their organization, such as a data privacy role, a single person who attends to safety will not be enough in an age of AI. While these existing risk management roles should continue, additional roles that address representation in training data, measure and mitigate algorithmic bias, rectify errors and misinformation in outputs, and tackle other concerns are also needed. Specific risks will emerge in particular components (such as a Large Language Model component vs. a Computer Vision component), in interactions between components, and through the different stages that take place while building and integrating AI systems (such as in collecting data sets for training, building AI models based on the data, and designing user interaction based on those models). Educational developers can examine how their internal roles and processes are organized to achieve consistent attention to responsible AI, and many are already doing this (Box F).

Box F: Examples of Developers' Public Documents about Responsibility

- DuoLingo released its Responsible AI Standards for its English Language Test, covering its approach to validity and reliability, fairness, privacy and security, and accountability and transparency. DuoLingo invited public comment on its document.
- In a document titled "What is Khan Academy's approach to responsible AI development?," the developer of the Khanmigo tutor discussed how it discloses risks, evaluates and mitigates risks, limits access to its software, and educates its developers about ethics.
- Instructure released an AI Governance Policy that discussed responsible AI use, transparency and accountability, bias and fairness, human-AI collaboration, training and education, and privacy.
- Grammarly has shared its Commitment to Responsible Innovation and Development of AI, which describes how it acts on commitments, e.g., "All models undergo bias and fairness evaluations. Our team of analytical linguists apply research and expertise to minimize bias and apply user feedback."
- Through work with many developer organizations, the Software and Information Industry Association produced seven Principles for AI in Education.

Many other development organizations have produced their own documents about their responsibility approach. In addition, as previously mentioned, eight major developers of AI models made voluntary commitments to manage risks (see White House Fact Sheet).

The Department observes that many educational and general AI developers are producing their own responsible AI principles (see Box F and the quote below). The public will benefit both from public commitment to responsibility and from accountability to such pledges. The Department therefore strongly urges developers to address how their "innovation stack" (components that work together to deliver new capabilities) can be integrated with a parallel "responsibility stack" (roles that work together to mitigate risks in every component and stage of development).

> *"Driven by a strong belief in purpose-built, beneficial AI, we prioritized domain specificity and safety from the outset. Through our Responsible AI initiative, we invite a diversity of voices to co-create solutions with us."*
>
> *— Sharad Sundararajan, Co-founder CDO/CIO, Merlyn Mind*

A Pathway from Designing for Education to Earning Trust

Through the listening sessions, the Department heard a wide range of specific concerns from developers and educators about AI in education.

Figure 2: A Development Process that Leads Toward Earning Trust

We categorized these concerns and organized this guide around five key overarching areas of shared responsibility, as follows (see also Figure 2):

1. **Designing for Education** recognizes that developers should begin by understanding values specific to education. Across many examples, the Department sees educators stepping up to articulate values such as centering humans in the loop and attending to priority educational challenges like reading, science, math, and computer science education. In addition, educator and student feedback should be incorporated into all aspects of product development, testing, and refinement to ensure student needs are fully addressed.

2. **Providing Evidence of Rationale and Impact** is important to making decisions about which edtech products to adopt or procure, especially where the goal of the product is to improve student outcomes. Both the Elementary and Secondary Education Act of 1965 (ESEA) and educational decision makers call for developers to provide evidence that products or services improve student outcomes. In procurement, for example, educational institutions are making clear demands for the evidence they require.

3. **Advancing Equity and Protecting Civil Rights** is an essential commitment of the Department and the Administration and a centrally important concern of constituents in the public listening sessions, both among developers and educators. Developers should be vigilant, for example, about issues of representation and bias in data sets, algorithmic discrimination in systems, and ensuring accessibility for individuals with disabilities.

4. **Ensuring Safety and Security** are emphasized in the Executive Order on AI and related Administration guidance. Educational decision makers are articulating their data privacy and security requirements with clarity and elaborating additional requirements—such as civil liberties—in an age of AI. To participate responsibly in the ecosystem, developers will need to detail the actions they will take to ensure the safety and security of users of AI.

5. **Promoting Transparency and Earning Trust** is an important overarching goal. Earning trust requires attending to all the values above and, in addition, has an important communication dimension that goes beyond output. For example, trust requires transparency and other public commitments to building mutual confidence among technology suppliers and users. Mutual engagement in defining and acting collaboratively among developers, educators, and other constituents builds trust.

Recommendation 1. Designing for Teaching and Learning

Developers of AI-enabled products and services should start with strong attention to education-specific values and visions, which are articulated in a mix of federal, state, and local resources, along with resources to support the use of AI in education produced by nonprofits and industry associations. Attending to ethics is an essential area of shared responsibility.

Key Ideas

- Resources developed by governments (federal, state, and local), nonprofits, and industry associations can provide developers with a good starting point for anchoring their work in educational values and visions and avoiding negative outcomes.
- Human-centered and humans-in-the-loop approaches that proactively include educators are emerging as key values that educational decision makers are demanding, and educator and student feedback should be incorporated throughout the development and testing process.
- Developers should attend to key ethical concepts such as transparency, justice and fairness, non-discrimination, non-maleficence/beneficence, privacy, pedagogical appropriateness, students' and teachers' rights, and well-being.
- Human factors, and human-centered design, and longstanding software development practices can provide a starting point for developers.

What to Know

The 2024 National Educational Technology Plan (NETP) presents a forward-thinking approach to reframing and realizing the potential of edtech to enhance the instructional core, reduce achievement gaps, and improve student learning in our schools. These are three critical national priorities. During an event to release the plan, U.S. Secretary of Education Miguel Cardona stated, "As we work to Raise the Bar in education, it's essential we focus on empowering teachers to become designers of active learning, using technology in effective ways to engage and inspire students." Inside the 2024 NETP, developers will find a wealth of information regarding educational visions, goals, and values related to the use of technology.

Notably, the 2024 NETP is not directly about AI. That is because a valid educational purpose and important unmet need should be the starting point for development, not excitement about what a particular technology can do.

At the federal level, the Department's 2023 AI Report provides additional guidance on aligning generative AI to what educators care about and need. The report has sections about student learning, supporting teachers, and improvement assessments, with many examples regarding

how AI could lead to advances in these areas. The Department has observed that the following two recommendations in the Report resonate consistently with educators in state, local, and international forums: (a) emphasize humans in the loop, and (b) align AI models to a shared vision for education. These provide important shared touchstones with educators as developers begin their design process with AI-enabled technologies.

State (Box C) and local resources provide additional guidance about AI in education. These resources partially anchor visions for AI in education on changing expectations regarding career readiness preparation for students' future employment. They highlight that developing teacher and student AI literacy is essential to responsible use of AI. Many of the resources provide advice on improving equity and inclusion. Data privacy, security, and safety are key risk areas that are important for developers to address and school leaders to manage. Overall, the resources provide developers with a better understanding of how educators are conceptualizing opportunities. In Box G, we list some opportunities that are frequently featured across these resources.

Box G: Opportunities for using AI in education

The Department sees many opportunities for improving academic outcomes, as well as accessibility and inclusion of students in academic programs. Here are some examples:

- Improving academic outcomes, accessibility, and inclusion for children and students with disabilities
- Providing students with more and better feedback and guidance as they learn curricular subject matter
- Addressing learner variability (including both access and inclusion) in all its aspects by better matching learning resources to each individual student's strengths and needs, as well as addressing the needs of historically underserved student populations
- Saving administrative time for teachers and enabling teachers to focus on their students
- Enabling teachers to incorporate research-based pedagogical principles, like those found in the What Works Clearinghouse practice guides, into their instructional plans
- Improving teacher professional learning by including opportunities to practice specific pedagogical strategies with simulated classrooms and students
- Reducing the cost to customize learning resources to build on strengths and interests of students in the locale of various educational communities
- Achieving efficiencies in school operations, such as school bus schedules

Of course, developers may surface equally important opportunities through their own engagements with educators. In the case of generative AI, the capabilities are still too new to be certain about where the best opportunities lie. Developers can strengthen alignment and validation of purpose by establishing strong feedback loops with educational communities at

every stage of product design, development, deployment, and refinement. When seeking feedback, it is important not only to include those in the most influential roles (e.g., superintendent, technology director, and/or other decision-maker), but also those who will be most affected by the educational product or service (e.g., classroom teacher, special education teacher, student, and family). Hearing diverse input is one way for developers to keep humans in the loop. Co-design with educators, a recommendation in the Department's 2023 AI Report, is a strong way to enact shared responsibility. However, as previously noted, involving educators in both designing and monitoring the use of AI is not a panacea and can impose responsibilities on educators that developers should rightfully own.

Developers frequently raised ethics as an important area of focus during the Department's listening sessions. Researchers and educators have been working together to develop ethical guidelines, and the Department expects this work to continue. (See the "Resources" section of this recommendation, for example.) A review of major ethics frameworks for AI in education found that the following general ethics concepts were applicable to education: transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom, and autonomy.

In addition to adapting these principles to more closely fit education settings, the review identified four additional principles specific to education: pedagogical appropriateness, children's rights, AI literacy, and educator well-being. These ethical principles are core components of designing AI systems that interact with children. For example, is it ethical to present an animated coach in a product as a humanoid persona, blurring the line between people and algorithms? Or, if AI is included in a system to support an aspect of teachers' well-being, what is the standard of care, and when should human caregivers be involved? Additional examples of ethical considerations in the design and implementation of AI-enabled tools will become clear as the use of these tools in education expands. (Figure 3). Regarding AI Literacy, the review states, "AI literacy underlines the educational importance of children and youth learning about AI so that they may become critically informed, as well as the need to build teachers' professional knowledge and parent awareness of AI."

Figure 3: A Synthesis of Ethics Themes

SAMPLE OF VALUE-CENTERED DESIGN PRINCIPLES FOR ETHICS IN EDUCATION	
General Ethics Themes	**Education Ethics Themes**
• Transparency • Justice and Fairness • Non-maleficence • Responsibility • Privacy • Beneficence • Freedom and Autonomy	• Pedagogical Appropriateness • Children's Rights • AI Literacy • Teacher Well-Being • Responsiveness to Student Needs

In February 2024, researchers at NIST suggested that building on long-standing concepts in the 1979 Belmont Report (which established principles for the protection of human subjects in research)—beneficence, respect for persons, and justice—can organize how developers address ethics in the age of AI. For example, beneficence can be established by collecting and sharing evidence that a product delivers expected benefits and by mitigating any situations in which individual users might experience harm (even though effects on average are positive). This applies across the life cycle of a solution, from prototyping to productizing. Additional ethics concepts can likewise be translated into practical steps that developers could take (and build on with measures many educational developers already routinely practice).

Likewise, developers should attend to a foundation of rights and respect for human dignity as they create new AI-enabled applications. This is consistent with language found in the Administration's Executive Order on AI:

> *"The interests of Americans who increasingly use, interact with, or purchase AI and AI-enabled products in their daily lives must be protected. . . . [which requires] appropriate safeguards against fraud, unintended bias, discrimination, infringements on privacy, and other harms from AI."*
>
> *— Executive Order on AI*

International resources such as The European Commission's Proposal for a Regulation of the European Parliament and of the Council: Laying Down Harmonised Rules on Artificial Intelligence put human dignity at the foreground of ethical considerations.

Product leads and their teams should not only be aware of ethical concerns but should also find ways in which ethics can be interwoven throughout the life cycle of product development. This applies from the initial ideation and prototyping stages and product deployment and continues in perpetuity as the solution improves—both autonomously and with human intervention—across the product's life cycle. Value-centered design is one approach that weaves people, purpose, and ethics together (see Resources). Also, the Association of Computing Machinery has useful guidance on ethics for developers (also in Resources).

Many organizations are already paying attention to human factors in their development approaches. AI may be relatively new to education, but simple to sophisticated examples have been prevalent for decades in other fields such as manufacturing, aviation, and retail. Further, the concept of human-centered design has a long history in computer science (see Resources section below) and is something educational developers already incorporate into their processes to varying degrees. In listening sessions, educators strongly advocated for their involvement not only in initial design but also in the development process, both to improve the system and to participate in explaining the use of AI in the system to other educators. Similarly, educational researchers have been cultivating space for youth voices throughout the development process. Most generally, it is important to include not only the users who are powerful but also those who will be most affected by an application of AI in education.

"As a LatinX and neurodiverse student, I think it's especially important to mitigate the systemic biases that are entrenched within AI models. Assembling a diverse team of people who specifically work to tackle this issue is essential to building safe AI."

—Nicholas Gertler, student and chair of AI Issue Advisory Council and AI & Education Advisor, Encode Justice

However, in some cases, human-centered design may only receive attention in the user experience (UX) layer of development or in the work of one specific team or department in an edtech company. For example, giving feedback to a student might be considered to be a UX feature. However, giving appropriate feedback to a student may depend on the quality of the product's records of a student's past learning, earlier pedagogical interventions, and what kinds of feedback work well for that student. Thus, the database aspects of giving high quality feedback are equally important to the UX aspects of giving high quality feedback. For an AI product or service to be human-centered, developers will likely need to incorporate human-centered methods throughout the layers of system design and deployment. Human-centered development of AI for education should occur throughout a developer's responsibility stack.

Questions to Ask

1. What can we learn from how written educational strategies (such as the NETP) and our educational customers describe the most important and equitable purposes for using AI?
2. How can our work align to the "humans-in-the-loop" recommendation in the Department's 2023 AI Report?
3. Through what feedback loops are we continually learning more from our users about how to align to educational purposes to meet the needs of diverse students and to respect the role of educators?
4. How does our team understand ethical considerations for AI in education, and what steps can we take to integrate ethics into our work?
5. How have children and youth, families, and educators from underserved settings been consulted and involved in design decisions?
6. How can we continually strengthen development disciplines (such as human factors and human-centered design) to address emerging AI-related features and risks?

Directions to Pursue

- Developers should familiarize themselves with relevant resources that express educational visions and strategies, including resources that are available at the national, state, and local levels as well as internationally.

- Developers should deepen their understanding of historical discrepancies in opportunity to learn and in learning outcomes for diverse student populations and how their products could contribute to equity for all students.
- Developers should be intentional about strengthening their feedback loops with educational user communities throughout the product life cycle, from defining the product's purpose to refining how it operates.
- Developers should engage ethics experts to guide their work and build their team's understanding of ethical issues in the day-to-day work of developing, deploying, and continually improving a product.
- Developers should involve educators and youth throughout the product development process, seeking to include not only those with power but also those who will be most affected by design choices in a product or service.
- Developers should explore obstacles to human judgment (e.g., a tendency to defer to the suggestions coming from a technology), to misunderstand limitations of AI-based inferences, or to underappreciate the potential for more risks to emerge once AI is deployed in educational settings and expand their understanding of human factors to encompass all the ways an AI-enabled system may enable or impede sound instructional and educational decisions.

Resources

- Batya Friedman & David G. Hendry's Value Sensitive Design: Shaping Technology with Moral Imagination
- Center for Humane Technology's Potential Policy Reforms Toolkit
- Center for Democracy & Technology's AI Policy Tracker
- NIST's Human-Centered Design Principles
- Organisation for Economic Co-operation and Development (OECD) & Education International's Opportunities, guidelines and guardrails for effective and equitable use of AI in education
- United Nations Educational, Scientific and Cultural Organization's (UNESCO's) Artificial Intelligence and the Futures of Learning project
- U.S. General Services Administration's Human-Centered Design Guide Series

Recommendation 2. Providing Evidence for Rationale and Impact

By clearly articulating how they have incorporated evidence-based practices into their products and how they intend to build new evidence about a product's usability and efficacy, developers and deployers of AI systems in education can work together toward responsible use. Developers of AI-enabled products and services share responsibility to explain how research informs the rationale (or logic) of their offering, to document and analyze data to make improvements and address risks, and to evaluate the impact on educators and students, especially those in historically underserved groups or settings.

Key Ideas

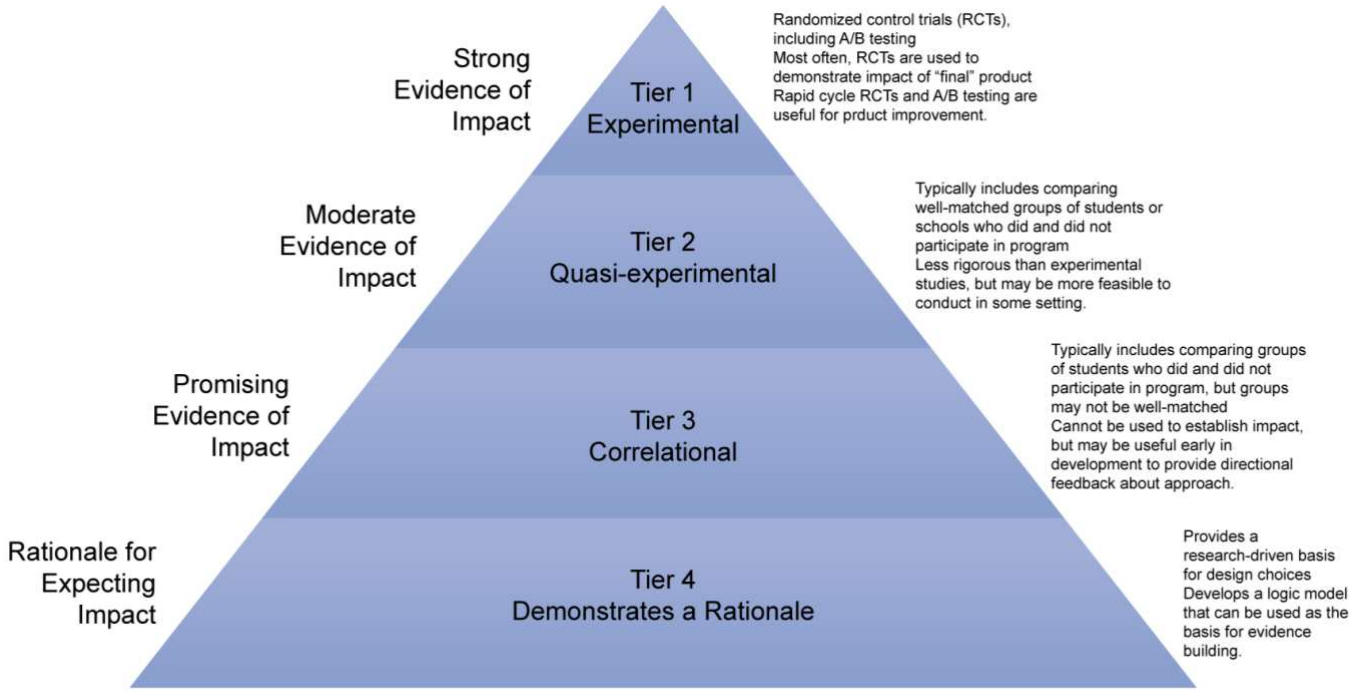
- The ESEA specifies four tiers of evidence (see ESEA section 8101(21)) that characterize the quality of research evidence that establishes whether an educational product or service has been shown to improve student outcomes. (See also the Department's September 2023 non-regulatory guidance.) Educational leaders increasingly demand such evidence when making adoption and procurement decisions.
- Developers should clearly articulate how existing evidence-based practices inform the design of their product or service. When specific evidence is not yet available for potentially innovative features, developers should be clear about the more general scientific rationale that justifies including the features in a product's design.
- Developers should be clear about the student outcomes their product or service is meant to improve. When the product or service provider seeks to measure those outcomes, the measure used should be high-quality (i.e., demonstrate acceptable levels of validity and reliability for the students and settings in which the product or service will be used).
- Developers should seek to build evidence about the risks and outcomes associated with, and impacts of, the use of their product and services. The evaluation process should be framed to measure to type of outcome the product targets and should take into account the importance of identifying and mitigating potential risks, especially risks that might differentially impact vulnerable and underserved populations. Undertaking the most rigorous study (e.g., a randomized controlled trial) may at first seem daunting. If this is the case, developers may first use less rigorous forms of evidence building (e.g., correlational studies) to describe how student demographic information and product use is associated with observed student outcomes. Over time, developers may proceed to more rigorous evaluations that support statements about a product's causal impact on student outcomes, such as by using randomized controlled trials or other rigorous approaches to evidence building.
- When making adoption decisions, states, districts, and schools typically consider the extent to which the evidence supporting a given product or service meaningfully includes the students and settings they seek to serve. As such, when building evidence about a program or service's relationship to, or impact on, student outcomes, developers should disaggregate their findings to be clear about "what works, for whom, and under what conditions." Considerations of "for whom" should address vulnerable and historically underserved student populations; for example, an AI-enabled product for mathematics teaching and learning should address the types of disparities in mathematics achievement documented via the National Assessment for Educational Progress via policy-relevant demographic variables.

What to Know

In 2015, the Every Student Succeeds Act reauthorized the ESEA. The ESEA encourages school and district decision makers to choose educational products, services, and interventions that have been shown to improve student outcomes through high-quality research and evaluation or that at least demonstrate a rationale that they will be effective. The Department also provides a document titled Non-Regulatory Guidance: Using Evidence to Strengthen Education Investments, which includes the evidence framework and definitions in the Education Department General Administrative Regulations. In addition, the Office of Educational Technology offers an Edtech Evidence Toolkit with resources including one-pagers, case studies, and examples. Developers of AI-enabled tools and platforms should know that required characteristics of educational evidence are specified in law and used in practice when educational leaders make adoption and procurement decisions and, with this knowledge, specifically define the outcomes their solutions can provide and how to measure them as evidence.

The ESEA defines four tiers of evidence, which are summarized in the Department's What Works Clearinghouse as shown below in Figure 4.

Figure 4: Tiers of Evidence



Early-stage developers may start by using evidence, including scientific theories, to develop the rationale for how their product or service is intended to strengthen student learning. The National Academies of Sciences, Engineering, and Medicine 2018 report "How People Learn II: Learners, Contexts, and Cultures" recommends assessing the degree to which learning environments are learner-centered, knowledge-centered, assessment-centered, and community-centered. Here are some examples of how to do this:

- Learner-Centered: How can an AI-enabled product support learners to be more active and collaborative as they strive to make sense of new information?
- Knowledge-Centered: How can an AI-enabled product activate a learner's prior knowledge and engage them in processes that actively strengthen understanding like elaborating, explaining, or critiquing?
- Assessment-Centered: How can an AI-enabled product provide students with more timely, relevant, and useful guidance and feedback when they encounter a difficulty?
- Community-Centered: How can an AI-enabled product support social interactions in which a student's peers, teachers, and other community members actively support an individual's strengths and needs?

"How People Learn II" also provides a list of five effective strategies to support learning:

- Retrieval practice
- Spaced practice
- Interleaved and varied practice
- Summarizing and drawing
- Explanations: elaborative interrogation, self-explanation, and teaching

There are many additional principles specific to age levels or subject domains, such as the science of reading, learning mathematics, or pursuing science and engineering. The Department's Institute of Education Sciences (IES), through its What Works Clearinghouse, provides general as well as domain-specific recommendations for evidence-based practices through its series of Practice Guides. Working with a teacher professional association dedicated to target school content is also often a good way to uncover learning principles specific to a domain. Another way is to contact scholarly societies or associations seeking a connection to a researcher active in the area the developer will pursue.

To hone their own rationale, which might be based on the above or additional modern learning principles, developers may seek feedback from pilot group users to inform design and development adjustments to the solution to improve efficacy. Evidence should inform both how a narrow product component has been designed (e.g., the ways in which mathematics is represented in symbols and graphics) and provide a scientific rationale for the foundational logic of their approach to improving learning. Foundational logic is sometimes called the

"theory of action" and is understood to be the integrated set of mechanisms that together lead to improvements to student outcomes (see Resources for more information).

Researchers and developers often follow the best practice of documenting their theory of action through a logic model. A logic model traces the connections from inputs (e.g., that a product or service provides) to teaching and learning processes (e.g., that teachers and students enact with support from the product) to outcomes (e.g., increased student achievement in a particular subject). Within a logic model, a major point of focus for developers, as they begin a journey toward strong evidence, should be defining a clear problem that is informed by robust root cause analysis, identifying an appropriate set of target outcomes, designing products or features that are likely to achieve those outcomes, and designing processes to measure those outcomes.

After establishing that their product or service "demonstrates a rationale" for use, developers may wish to consider how they can build evidence that it is effective. The strength of evidence supporting claims about a product or service's effectiveness can vary from relatively lower (Tier 3 or Promising Evidence) to relatively higher (Tier 1 or Strong Evidence). Lower-quality evidence (e.g., Tier 3) offers less confidence about effectiveness and speaks only to an *association* or *relationship* between a product or service and student outcomes. Higher-quality evidence (e.g., Tier 1) offers more confidence about effectiveness, leading to conclusions that a product or service *caused* a student outcome.

As shown in Figure 4, the primary factor that distinguishes lower-quality evidence from higher-quality evidence is the type of research that is used to evaluate a product or service's effectiveness. Not all developers will have experience in designing and conducting high-quality research and, as such, will seek out the assistance of outside experts. Notably, even developers with research capacity may elect to partner with an independent evaluator to lend additional credence to their effectiveness findings.

Successful evidence-building at any tier depends on both robust forethought and adequate resources. It is almost always easier to design a program or service from the beginning with the goal of high-quality evidence building in mind than to add that goal when a program or service is nearing completion or final. As such, developers should consult experts in evaluation early in the design (or improvement) process. Although evidence-building can occur with lower costs when thoughtfully incorporated into a new or redesigned program, additional financial resources supporting evaluation—particularly an external evaluator—may be needed. Grant funding for research is available in multiple federal programs (see Resources) and through some philanthropic foundations.

Measurement of educational outcomes is always imperfect and always a focus of improvement for the field. The Department's 2023 AI Report includes a section on formative assessment which emphasizes measuring what matters; specifically, that educators call for broader ways of capturing what students know and can do (compared to traditional multiple-choice assessments, for example). Evidence from standardized end-of-course measures (culminating or summative assessment) will continue to play an important role in assessing the efficacy of policies, systems,

and practices that support student learning. However, formative and diagnostic evidence collected using additional relevant measures (continuous or formative assessment) also provides valuable information for educators, parents, and students about learning and academic performance. This sort of evidence also provides faster feedback cycles for all engaged parties. Developers may wish to engage with outside experts to learn more about how to connect their vision to measurable outcomes and how to identify measures with strong psychometric qualities (e.g., high levels of validity and reliability).

The tiers of evidence describe rationale and impact but leave an additional area of evidence open: how developers document risks and safeguards in their products by collecting data, analyzing what is working and what needs change, and guiding improvements. It is a common practice for many educational developers to provide case studies of how they worked with schools or districts to field pilot and test their product. They may use A/B testing to identify the features of their product that support needs across a variety of educational settings.

In terms of shared responsibility, educational decision-makers also value transparency for the means employed by the solution developer to field test and improve their product, as well as what product certifications have been secured through evaluations by trusted third-party learning organizations working in edtech. With regard to continuous improvement, some developers are exploring how to streamline the process of engaging with external researchers who could test improvements to their platform (see the Institute of Education Sciences' funded SEERNet, for example), while safeguarding privacy and other considerations. For example, developers are working with researchers to validate in-platform measures that can be used to study learning outcomes while reducing data collection costs. Developers can also incorporate interfaces into their platform that enable researchers to specify how to deliver variable resources to students while automatically collecting data—without requiring researchers to know the details of the developer's platform code. Developers can also standardize mechanisms for protecting the privacy of student data to enable researchers to perform analyses without enabling researchers to identify specific students.

Finally, evaluations are typically more valuable to decision makers when they address "for whom and under what conditions" because schools and districts want to know if the results are likely to generalize to their population and setting. With such a dynamic range of diversity in our nation's schools and districts, an approach that is only tested with students in a large metropolitan school district in one state may not work optimally for staff and students in a small rural school or a district in another state. Likewise, other questions arise, such as whether children and students with disabilities were included in the study population, and were there any noteworthy differential impacts for these students? Was the product effective for both students with low and high prior performance or only for students with average performance? Does the product work well for multilingual learners? There are many other relevant questions about "for whom and under what conditions." Developers can purposely vary the populations and settings in which they evaluate their offering to accumulate evidence that respects variability among learners and across settings.

This approach to addressing variability between context and learners has significant relevance for products that use AI to adapt to students' strengths and needs, as products may seek to strongly serve a greater diversity of students or function in a wider range of conditions. That AI technologies avoid harm and produce "the greatest good for the most people" cannot be taken for granted and must be specifically investigated. (An important and related concept appears in the equity-centered section; developers should seek to demonstrate their process for developing AI algorithms and models to minimize unfair bias.) The Department acknowledges it is expensive and time-consuming to study all possible variables. Developers should weigh the potential for harm and the level of confidence that a risk has been mitigated. Schools should protect students from harm, for example, regarding infringing students' civil rights—and thus algorithms that unfairly impact students' broad opportunities to learn and advance could be high risk and high priority for strong evidence. An example of a lower risk application might be a productivity tool that helps teachers do routine aspects of their work, such as configuring classroom edtech for a particular activity that the teacher has planned, and which can easily be supervised and corrected by the teacher. High-quality evaluation studies, similarly, can detect not only whether an educational resource benefits students on average but also whether it is ineffective for groups of students.

Questions To Ask

1. How could our team fulfill the evidence-based rationale or Tier 4 level and continue to expand the evidence base for diverse populations?

2. How could our product rationale build on or align with existing bodies of evidence or theory about how people learn or teach effectively, especially including students in historically underserved populations?

3. How (including with what partners) can we begin collecting evidence to demonstrate the potential for positive impacts for a diversity of students related to our use of AI in education?

4. How can we ensure that educational decision makers have access to the evidence and other information to make responsible choices about using AI applications?

5. What is our long-term plan to generate rigorous evidence, including for whom and under what conditions our application works and for expanding the populations for whom the application is effective?

Directions to Pursue

- Developers should ask potential customers how they use specific kinds of evidence in their decision-making process and "what success would look like," which would go beyond asking potential customers to list relevant local, state, or other procurement requirements.

- Developers should seek to form partnerships with researchers early in their design work so that modern learning principles can be employed to maximum advantage.
- Developers should seek to form partnerships with educators and users to conduct field tests throughout the life cycle of their product.
- Developers should involve those who will be most affected by a product in collecting and interpreting evidence.
- Developers should collect evidence not only on efficacy, but also related to safety, security, trust, and other issues.
- Developers should collate the workstreams from above into cogent and living documentation that is regularly updated and housed publicly online for the sake of transparency.

Resources

- Relevant funding sources
    - IES Small Business Innovation Research Program
    - IES Education Research and Special Education Research Grant Programs
    - The Department's Education and Innovation Research Program
    - The National Science Foundation's Division of Research of Learning
    - Many philanthropies also provide funds for educators, researchers, and developers to work together, with shared responsibility for evidence. The Bill and Melinda Gates Foundation's AIMS Collaboratory is one example.
- Resources on methods for building evidence
    - The Institute of Education Science Resources
        - The Logic Model Workshop Toolkit
        - IES Standards for Excellence in Education Research
        - IES/National Science Foundation Common Guidelines
        - Companion Guidelines on Replication & Reproducibility in Education Research
        - What Works Clearinghouse Find What Works Guides and Practice Guides
    - The Office of Educational Technology's EdTech Evidence Toolkit

    - The U.S. Department of Health and Human Services' Rapid Cycle Evaluations
- Organizations that provide information about evidence building
    - LearnPlatform
    - LeanLab
    - Common Sense
    - International Society for Technology in Education (ISTE)
    - Digital Promise
    - EdTech Evidence Exchange
    - International Certification of Evidence of Impact in Education

Recommendation 3. Advancing Equity and Protecting Civil Rights

Developers and educators share responsibility for advancing equity and protecting students' civil rights. Developers who are equity-centered will better address the potential for algorithmic discrimination, guard against civil rights violations, advance accessibility for all users, especially children and students with disabilities, and move to close overall gaps in design, use, and access of edtech.

Key Ideas

- Algorithmic discrimination could result in unfair distribution of opportunities to learn, resources and supports for learning, or outcomes of learning. NIST has identified three major categories of AI bias to be considered and managed: systemic, computational, and human, all of which can occur in the absence of prejudice, partiality, or discriminatory intent.
- Civil rights in educational settings are established in law and enforced by the Department's Office for Civil Rights. Developers should be well-informed about existing civil rights laws that apply to educational settings and design to comply with these laws. Existing civil rights laws apply no matter to what degree AI is implicated in the violation.
- Pre-existing and forthcoming AI training data sets should seek to reduce bias and represent educational user diversity. Educators are already expressing high awareness of these potential issues.
- Inclusion and accessibility are areas in which the capabilities of AI to support multiple forms of human interaction and augment human strengths and needs may be particularly beneficial.
- Digital equity encompasses attention to gaps in design, use, and access.

What to Know

Advancing equity is a broad concept but also points to a set of more specific considerations that developers should center in their work. This guide names civil rights[4], algorithmic discrimination, and accessibility as specific equity-related topics and uses the term "digital equity[5]" to point to additional issues (i.e., fairness in the affordability of technologies) that are important yet may not rise to the level of violations of laws. Existing laws (including civil rights laws) are paramount and apply to situations where any variation of AI leads to any discrimination across the continuum of a child's learning experience. Civil rights laws protect students against discrimination based on protected characteristics[6]— and apply to learning experiences inside and outside the classroom during the school day.

The Executive Order on AI directs many federal agencies with advancing the connection of AI equity and civil rights, both separately and in coordination across agencies. Developers should monitor agency websites to seek to stay abreast of policy guidance and other resources advanced by the Department and by other agencies, such as NIST and the U.S. Department of Justice, related to advancing equity and protecting civil rights.

Further, the Executive Order on AI directs the Assistant Attorney General in charge of the Civil Rights Division to meet with all federal civil rights offices "to discuss comprehensive use of their respective authorities and offices to prevent and address discrimination in the use of automated systems." This meeting occurred on January 10, 2024 (see Readout). The Department's Office for Civil Rights may evaluate and/or investigate allegations of civil rights violations stemming from the use of AI-enabled systems in educational settings.

Algorithmic Discrimination

Likewise, the Biden-Harris Administration made its position on the potential for algorithmic discrimination abundantly clear in the Executive Order on AI:

> *"My Administration cannot—and will not—tolerate the use of AI to disadvantage those who are already too often denied equal opportunity and justice."*

The Office of Science and Technology's Blueprint for an AI Bill of Rights defines "algorithmic discrimination" and outlines what developers should do as follows:

[4] As appropriate, developers should consider the civil rights of educators, organizational employees, or other end users. Developers may find this overview of the Department of Justice's Civil Rights Division's work on AI and civil rights of interest in making such determinations. Importantly, this guide does not attempt to discuss all of the civil rights provisions that could potentially apply to AI use in education settings.

[5] Section 60302(10) of the Infrastructure Investment and Jobs Act defines "digital equity" as "the condition in which individuals and communities have the information technology capacity that is needed for full participation in the society and economy of the United States."

[6] Title VI of the Civil Rights Act of 1964, as amended, 42 U.S.C. § 2000d, 34 C.F.R. Part 100; Title IX of the Education Amendments of 1972, as amended, 20 U.S.C. § 1681, et seq., 34 C.F.R. Part 106; Section 504 of the Rehabilitation Act of 1973, as amended, 29 U.S.C. § 794, 34 C.F.R. Part 104; The Americans with Disabilities Act of 1990, as amended, 42 U.S.C. §§ 12131, et seq., 28 C.F.R. Parts 35 and 36; The Age Discrimination Act of 1975, as amended, 42 U.S.C. § 6101, et seq., 34 C.F.R. Part 110.

> *Algorithmic discrimination occurs when automated systems contribute to unjustified different treatment or impacts disfavoring people based on their race, color, ethnicity, sex (including pregnancy, childbirth, and related medical conditions, gender identity, intersex status, and sexual orientation), religion, age, national origin, disability, veteran status, genetic information, or any other classification protected by law. Depending on the specific circumstances, such algorithmic discrimination may violate legal protections. Designers, developers, and deployers of automated systems should take proactive and continuous measures to protect individuals and communities from algorithmic discrimination and to use and design systems in an equitable way.*

Developers should proactively and continuously test AI products or services in education to mitigate the risk of algorithmic discrimination. In addition, training educators on both proper and inappropriate use of their solutions may mitigate issues of algorithmic discrimination in specific educational applications of AI (or prior uses of machine learning) according to researchers. Outcomes of algorithmic use of such data in education could deny students in a protected class from equitable opportunities in learning and achievement.

What is important is that the overall impact should not be inequitable. This applies to the use of AI enabled solutions in curriculum, as well as any technology or solution that can be used for monitoring behavior, classroom management, or discipline. Educators have many specific bias concerns related to computer vision algorithms and yet recognize that other forms of input such as speech recognition could also prove problematic.

The Department notes that the potential for algorithmic discrimination will not be limited to an application that makes obviously big decisions such as guiding student course or career selection but could also occur in a series of smaller decisions (for example, in the pacing or content of technology-based lessons), which in aggregate effect of the smaller decisions leads to an inequitable learning opportunity for students. Developers can address such concerns early in the process of building and training an AI, including by collecting representative data, exercising care in how data are curated and how algorithms are selected, testing for bias, and more.

Regarding shared responsibility, throughout the Department's listening sessions, educators consistently expressed strong concerns about bias in the data sets that are used to train AI models, both in foundation models and in tuning models to educational applications. Bias in the model's performance may occur because existing data sets inherently include historical biases and omissions—data sets may not be representative of the educational population who will participate as users. Developers should likewise build in opportunities for human review and strengthen features in products that increase transparency about the outputs that AI generates or the reasoning behind AI-based recommendations. They should also proactively design their products for learner variability, as will be discussed below.

"To simultaneously keep students safe and secure and train AI for equitable solutions with broad and diverse data is a defining challenge for this space. Simply saying 'no data use' will not keep all students safe or equip the equitable access we know all students deserve. As this field is rapidly evolving, thoughtful solutions must be identified to ensure safety, accountability, fairness, and effectiveness."

—Karl Rectanus, edtech entrepreneur

While risks exist, early research suggests[7],[8] that AI could be useful in helping better support children and students with disabilities, multilingual learners, and other populations that have long encountered barriers to learning with resources designed with less accommodation to their needs. The Department's 2023 AI Report outlines additional key areas of equity in the design, development, and deployment of AI-enabled systems in education with these examples:

- The report observes that adaptive algorithms in past edtech products often were more focused on deficits, weaknesses, mistakes, or gaps. Although addressing mistakes and errors is important in the development of AI-enabled tools, developers should consider balancing this approach in the development of such tools with a more asset-based design that deploys additional AI capabilities focused on assets to build on students' strengths and interests, also aligned with available community resources and assets.

- The capabilities of AI to support a wider range of inputs (e.g., speech, gesture, drawing) and outputs (e.g., translations among languages, annotations of images with language, automated production of American Sign Language) can provide additional supports to children and students with disabilities and can do so more evenly across varied learning activities and instructional resources as well as assessments.

- Similarly, the newer capabilities of AI can be aligned to achieve advances in equitable support for multilingual learners via translation support and identification of culturally responsive resources to accompany instruction.

  "We are excited about the potential of AI to allow us to adapt content more easily to students' linguistic needs and personal interests. For example, we have been field testing an AI-based system that prompts students for their interests and writes a word problem that fits their interests as well as the mathematics topic they are learning."

  —Dr. Steve Ritter, Founder and Chief Scientist at Carnegie Learning

More generally, the 2023 AI Report highlights how AI could be used to adapt instructional resources to all aspects of learner variability; whereas prior edtech services may have been most effective for students most similar to a developer's intended target population, developers could seek to use AI to serve a "long tail" (wider distribution) of student strengths and needs.

[7] https://journals.sagepub.com/doi/abs/10.1177/00400599241231237

[8] https://library.iated.org/view/MEHIGAN2024CON

Accessibility

As they seek to support learner variability, developers should be aware of and follow requirements of the Individuals with Disabilities Education Act (IDEA, as amended in 2004). Also developers should review Section 504 of the Rehabilitation Act of 1973 (Section 504). IDEA highlights the importance of educational resources that *leverage students' strengths* and not only resources that address their challenges or needs. Section 504 prohibits discrimination on the basis of disability by recipients of federal financial assistance. Digital accessibility is a component of accessibility for students with disabilities. Developers should look both broadly and narrowly for support toward incorporating digital accessibility into their solutions.

The Web Content Accessibility Guidelines (WCAG), developed and maintained by the World Wide Web Consortium (W3C), is an approved ISO standard and is recognized by developers as the benchmark for creating content that is accessible without limitations to all users. AI can enable edtech developers to support interactions with people using a broader range of modalities (e.g., speech, gesture, American Sign Language, etc.). Incorporation of such capabilities could help not only specific learner populations but all learners as well. More focused on education solution design, Universal Design for Learning (UDL) is one well-established framework for guiding design of tools that "improve and optimize teaching and learning for all people based on scientific insights into how humans learn." Developers may find UDL's three broad guidelines to be a good starting point for conceptualizing how AI could improve learning via their product or service:

1. **Engagement**. Presenting educational content in new interactive formats can increase student engagement. As an example, with appropriate guardrails, AI can support U.S. or World History courses by enabling students to interact with historical interviews of real people.

2. **Representation**. Presenting educational content options via multiple entry points of representations (e.g., text, audio, graphics, animation) benefits learners; in a simple example, AI may support generating more useful "alt text" to accompany images accessed by screen readers, even if humans remain in the loop to verify AI generated alt text.

3. **Action and Expression.** Giving students enhanced opportunities to act and express themselves improves learning; in one research-based example, students can learn science by engaging in an animated, interactive narrative where AI makes the plot, characters, and dialogue adaptive.

Digital Equity

More broadly, the Office of Educational Technology's 2022 report, Advancing Digital Equity for All, describes ongoing gaps in access to devices and broadband connectivity that impact educational opportunity but goes beyond access to also include affordability and unequal adoption as concerns, as well. We provide examples of potential pitfalls that developers should be aware of in developing and deploying AI systems:

- Access: AI-enabled applications may first appear in districts of more wealthy communities, or conversely, under-resourced schools may rely upon more affordable AI resources instead of human resources.

- Affordability: In more affluent households, guardians may be able to afford premium subscriptions for the most powerful versions—especially if students are expected to use AI-enabled tools at home.

- Adoption: Lack of community buy-in for valuable AI-enabled educational products may occur when communities are less informed or less involved in the design and marketing of those products or when other necessary resources (such as teacher professional learning and development and accessibility of tools) are differentially available.

Developers serving edtech interests have equity-related considerations other sectors may not; they should be aware of the potential pitfalls and the steps they can take toward equitable access, affordability, and adoption. The 2024 NETP has additional recommendations regarding equity that may provide useful guidance to developers. Specifically, it discusses three types of gaps:

1. Digital Use Divide: addressing opportunities to improve how students use technology to enhance their learning, including dynamic applications of technology to explore, create, and engage in critical analysis of academic content and knowledge; and ensuring that students have equitable opportunities to use technology for learning.

2. Digital Design Divide: addressing opportunities for educators to expand their professional learning and build the capacities necessary to design learning experiences enabled by technology that serve the diversity of their students.

3. Digital Access Divide: addressing opportunities for students and educators to gain equitable access to educational technology, including connectivity, devices, and digital content. This also includes accessibility and digital health, safety, and citizenship as key elements of digital access.

Questions To Ask

1. How can we connect civil rights and digital equity to specific ideas that inform our work as developers and ensure we are following applicable federal laws?

2. How does the role of the education leaders in protecting civil rights relate to our product or service?

3. What steps can we take to audit and remove the potential bias or algorithmic discrimination in our product, with special attention to mitigating any impacts for vulnerable or underserved populations?

4. How could we leverage AI in our product specifically to enhance accessibility and inclusion?

5. What is our long-term strategy to be a positive force for digital equity in all its dimensions, from design all the way through distribution and use?

Directions to Pursue

- Developers can infuse their organizational culture with equity and civil rights priorities from underlying data sets used for training algorithms to UI/UX choices.
- Developers may establish or improve a review process/checklist for new platforms, enhancements, and/or extensions to ensure broad representation in solution performance.
- Developers may build feedback loop mechanisms with organizations and relationships with expert practitioners for equitable learning experience design.
- Developers can work to stay current with both mainstream and edtech standards bodies and their pending work to address racism and other forms of algorithmic discrimination in AI-enabled products and services.
- Developers may participate in regular third-party review processes for eliminating bias from underlying databases, algorithms, and even UI/UX design elements that exclude certain cohorts from equitable user experiences.

Resources

- NIST is working on a standard for identifying bias in AI, with three main ways in which bias occurs (systemic, statistical/computational, and human), which intersect with different foci for mitigating bias (in datasets and models, during development of applications, and as applications of AI are used in the field).
- CAST's Universal Design for Learning is a research-based framework that can be used to guide applications of AI to build on students' strengths.
- The Center for Democracy and Technology provides extensive resources and guidance on equity, civil rights, and AI.
- Culturally Responsive Teaching and other design guidelines of processes are available to do equity-relevant design.
- The Leadership Conference works to ensure that new technologies further civil rights protections, and also works on related Digital Equity issues.
- The Learner Variability Navigator is a tool to find research-based strategies to support whole-child learning.
- Several nonprofits have design frameworks and services aimed at addressing equity. Two examples are the National Equity Project and Center for Inclusive Innovation.

Recommendation 4. Ensuring Safety and Security

Educational leaders and decision makers are looking to developers for partnership in strong plans to both address well-known risks and manage the broader range of potential risks so that AI technologies that improve teaching and learning can be implemented safely for students, staff, and communities. Developers who are safety- and security-centered will prioritize protecting students' and teachers' data security and privacy; developers will also acknowledge that risks now go beyond these broadly known, longstanding edtech risks and therefore conduct risk identification, prioritization, and management throughout their development and deployment processes.

Key Ideas

- Developers must be aware of the federal laws and related guidance regarding privacy and data security, such as these:
    - Family Educational Rights and Privacy Act (FERPA)
    - Protection of Pupil Rights Amendment (PPRA)
    - Children's Online Privacy Protection Act (COPPA)
    - Children's Internet Protection Act (CIPA)
- Protecting privacy and enhancing cybersecurity are key issues that school technology leaders manage as they procure, implement, and monitor technology in their educational institutions.
- Developers of AI applications for education must identify and mitigate risks that go beyond privacy and cybersecurity (see Box E).
- NIST has produced an AI Risk Management Framework that can guide a comprehensive and continuous process for identifying, prioritizing, and addressing risks.

What to Know

Privacy and data security are the aspects of edtech where the strongest guidelines and guardrails already exist. Most participants in the edtech marketplace have been actively addressing privacy and cybersecurity for many years before generative AI became widely available and will continue to require strong safeguards. The Center for Democracy and Technology reports privacy as a chief concern among parents and students using edtech while learning at school—especially for students with individualized education plans or 504 plans.

Educational leaders are committed to procuring tools that protect their students, staff, and communities. The Department provides support around federal requirements and best practices for protecting student privacy. For example, the Department's Student Privacy Policy Office (SPPO) offers extensive information on protecting student privacy in accordance with applicable federal laws and regulations. SPPO resources include information on the Family Educational Rights and Privacy Act (FERPA) and the Protection of Pupil Rights Amendment (PPRA), which apply to educational agencies and institutions receiving Department funds. The Department's Office of Educational Technology and the Privacy Technical Assistance Center (PTAC) host content and events to support developers in safeguarding student data. In addition, the Children's Online Privacy Protection Act (COPPA) adds protections for students younger than 13 years old, and the Children's Internet Protection Act (CIPA) addresses risks related to harmful content in schools and libraries. Developers must know these laws.

Developers should be aware that both states and associations provide leadership and direction to the work of leaders in local educational agencies (LEAs) as they tackle these issues. As one example, the Utah State Board of Education has a Student Data Privacy Team focused on supporting LEA efforts to protect student data. Most states have similar initiatives. The Consortium for School Networking (CoSN), an organization that supports school information technology professionals, has produced a parallel NIST cybersecurity framework for education to assist with classifying and understanding resources. Their membership and resources reflect strength in both cybersecurity and its extension to AI-enabled products and services, including a readiness checklist developed in conjunction with other learning organizations. The Data Quality Campaign provides extensive coverage of student data privacy and security policies, including information on how state and local jurisdictions go beyond federal protections. Additional resources are offered by the Student Data Privacy Consortium, which brings together constituent groups in the edtech marketplace to set expectations about student data privacy.

> *"AI isn't a monolith. When developers think about responsible design and use of AI tools in education, it's important to consider the range of AI applications and align their risk mitigation strategies to the relative risk of each tool. One strategy is to keep humans in the loop—for example, involving educators at optimal moments that help minimize the risks and maximize the value of these new technologies."*
>
> *—Teddy Hartman, Head of Privacy and Trust, Pear Deck Learning*

Risks of AI in educational settings extend beyond the well-known challenges of data security and privacy. The Department urges developers to rapidly begin to identify, prioritize, and manage additional risks. Consider the following examples, which use the categories in Box E and are based on publicly reported incidents:

- In an example of a race to release risk, generative AI chatbots became widely available to schools before any guidance was available to teachers and students on acceptable use, distracting educators from managing the core functions of teaching and learning.

- In an example of a Bias and Fairness risk, test proctoring systems that rely on facial recognition algorithms were suspected of unfairly and disproportionately disciplining non-white students.
- In an example of Harmful Content risks, students often use tools to create images for use in their school projects, and generative AI has been found to offer negative and hurtful stereotypes when given queries to construct an image for "black girls."
- In an example of a Malicious Use Risk, students have been found to use generative AI in cyberbullying, for example, constructing false and negative narratives about fellow students or false images about fellow students.
- In an example of Hallucination Risk and Wrong Information Risk, generative AI has been found to produce output that describes historical figures who never existed and to give wrong answers to math problems.
- In an example of Transparency Risks, educational procurement officers face pressure to allow various AI-enabled technology into schools, well before adequate information is available that would allow them to apply good judgment and ask hard questions about data sources, algorithms, risk mitigation, and other requirements.
- In an example of Underprepared User Risks, human decision makers sometimes defer to algorithms, which weakens the important role of educator judgment in how to support students and undermines the human in the loop recommendation from the Department's 2023 AI Report. Likewise, human decision makers sometimes absorb biases from the output of AI, interfering with their own better judgement.

An important starting point for mitigating AI-related risks is NIST's AI risk management framework. Although this framework is applicable to sectors broader than education, it can offer valuable guidance as developers identify and implement specific risk mitigation approaches that are more uniquely tailored to the education sector and encompass all identifiable risks. See Figure 6 and Figure 7.

Regarding **risk identification**, risks can occur at multiple scales, from harming individual people to harming an ecosystem. In addition, per the earlier discussion, AI-enabled systems for education will be developed with components from organizations outside the primary developer, and thus risks may emerge along a supply chain of components that are utilized in an overall product or service. Supply chain vulnerabilities can occur when a component is judged to be "safe" at an earlier time but then changes its behavior without notice to the component user; for example, foundational models have been observed to degrade at times for unknown reasons or to suddenly begin making errors as developers tweak their algorithms. Regarding **risk prioritization**, the European Union AI Act defines levels of risk. Currently, there is no similar official definition of levels of educational risks in the United States; however, developers are beginning to index specific risks in their applications and develop targeted

mitigation strategies. The Department urges developers, working in their own organizations and in shared responsibility with educators, to consider how to identify and prioritize risks.

> *"Our team developed a detailed list of potential risks as we designed our AI-enabled math tutoring product. Then we prioritized the risks and developed our responses. We'd love for educators to ask us about these specific risks, and we'd welcome feedback on how we've addressed them."*
>
> *—Kristen DiCerbo, Chief Learning Officer, Khan Academy*

Figure 6: Categories of Harm Described in NIST's AI Risk Management Framework

The Department, following the lead of NIST, suggests developers take a lifecycle approach to managing these and other risks that must be identified (Figure 7). The four aspects of a lifecycle risk management framework are:

1. **Govern**: Cultivating a culture of risk management within every edtech developer organization, for example, ensuring that developers in every phase or component of product development are aware of potential risks, their responsibilities to identify product-specific risks, and how to engage with product managers to mitigate risks.

2. **Map**: Recognizing the special features and challenges of school contexts and identifying specific risks that arise in those contexts. Developers can look throughout this document for examples of the specific kinds of risks that arise in school contexts as well as their special responsibilities and obligations under law when working with minors and students.

3. **Measure**: Documenting, analyzing, and tracking risks not only as the product is developed, but also as it is field tested in classrooms or other educational settings and more widely distributed and used. For example, developers can apply research-based methods to measure bias; the Department's 2023 AI Report included an extensive

discussion on how techniques from educational psychometrics could be helpful in ensuring fairness of AI-enabled educational resources or assessments.

4. **Manage**: Prioritizing risks and acting to protect students, teachers, and educational communities. Throughout listening sessions, the Department learned about development organizations that are prioritizing risks early in their development process, during the process of building each product feature or component, and after the product is launched.

Figure 7: Overview of the NIST AI Risk Management Framework

Educators look to developers to share responsibility for managing risks. On the one hand, state guidance resources (such as those mentioned above in Box C) instruct school districts to govern, map, measure, and manage risks. On the other hand, risk management capacity for AI in schools is likely limited and, thus, assumptions that risks known to a developer or known to the public can be managed by those closest to teaching and learning are unlikely to be borne out. Educational institutions will need partnerships with developers to help them manage risks. Shared risk management measures, implemented in customer service agreements between solution providers and end users in schools, can provide an added value that is not only ethical but also fosters trust through deliberate attention and thorough, regularly updated documentation to mitigate any potential risks, known or unknown. For example, PTAC offers a resource on developing a model terms of service.

Questions to Ask

1. What are we hearing as top safety and security concerns from educators regarding our product or similar products?
2. How do applicable federal, state, and local laws and guidance, such as the White House Blueprint for an AI Bill of Rights, apply to our product or service?
3. What specific privacy, security, or other safety risks are most relevant to our product and to vulnerable groups, and how can we mitigate them?
4. How can our organization systematically govern, map, measure, and manage AI risks beyond privacy and security?
5. What overall risk management strategy throughout our organization would lead to growing our positive reputation for delivering AI-enabled edtech with the strongest protections for privacy, security, and safety?

Directions to Pursue

- Developers should draft clear and plain language disclosures about how an organization protects student data security and privacy.
- Developers should bolster accountability efforts via audits or other procedures for inspecting and testing protections, as well as obtaining feedback on risks from end users, with special attention to vulnerable and underserved populations.
- Developers should collaborate across product lines or with other companies to articulate shared standards or approaches for addressing risks in educational products and services that incorporate AI, creating approaches to address issues of risk not only within an organization but also with upstream suppliers and downstream consumers.
- Developers should cultivate awareness of how the public, users, and regulators perceive levels of risks related to AI educational products and services and should respond if noteworthy risks emerge in an offering.
- Developers should prioritize staying abreast of rapidly evolving legislation and other governance activities related to AI in the federal government, in states, and locally.
- Developers should be thoughtful about addressing the interplay between state and federal policies. At the time of publication, more states are developing and releasing AI policies.

Resources

- The National Science and Technology Council's National Strategy to Advance Privacy-Preserving Data Sharing and Analytics.
- Cyber and Information Security Agency's Secure by Design Framework
- Software and Information Industry Association principles on developing AI-enabled edtech products
- Data Quality Campaign resources on protecting student privacy
- CoSN's NIST Cybersecurity Alignment for K-12.
- Additional resources at the Center for Democracy and Technology.

Developers who are working internationally may also wish to consider the European Union AI Act's levels of consequentiality of educational risks; more generally, staying abreast of developments worldwide may be important. One useful example is the Canton of Zurich's Legal Best Practices regarding AI in education.

Recommendation 5. Promoting Transparency and Earning Trust

As a sound business approach to the education market, developers engage in two-way communications with educators, students, and others in the ecosystem about AI, including promoting transparency for how AI has been implemented in an educational application, addressing concerns, and working together to expand the strength of shared responsibility.

Key Ideas

- Transparency contributes to trust.
- "Trustworthy systems" is an important technical area of research and development (e.g., the NIST AI Risk Management Framework), and trust is a relationship of mutual confidence among those who create AI-enabled educational systems and those who use them.
- Developers can grow trust by contributing to AI literacy among educators, parents and caregivers, students, and other constituents; conversely, without strong AI literacy, assurances by developers may fail to earn trust.
- To sustain trust, developers should tend to the level of ethics training among their teams.

What to Know

Developers who attended the Department's listening sessions viewed all four recommendations previously addressed in this guide (i.e., designing for education, evidence, equity and civil rights, and safety and security) as essential to developing trust in their education markets and communities they serve. Another way of representing this multifactor approach to earning trust is NIST's analysis of the characteristics of trustworthy AI systems. See Figure 8.

> *"Schools deserve AI solutions that are purpose-built for learning. That means solutions must adhere to the highest levels of safety, privacy, and security with models that are hallucination resistant and trained on vetted education-specific content. That's why we're building AI models specifically for education."*
>
> *—Latha Ramanan, Head of Responsible AI, Merlyn Mind*

Figure 8: Characteristics of Trustworthy AI Systems

The Department encourages developers to attend both to *trust* as the mutual confidence of two parties (e.g., relationships among developers and adopters of edtech) and *trustworthiness* as ascertainable properties of a technical system (as indicated in the image above).

An example based on the potential for AI to support "stealth assessment" may clarify the importance of transparency and relationships to trust. In its original meaning, "stealth" was a synonym for "unobtrusive" and was intended to provide authentic and supportive feedback to students while maintaining high student engagement in learning[9]. However, "stealth" can also imply surveillance—that students may be measured invisibly, without their knowledge, and with no direct and obvious benefit to their learning. With regard to the original meaning, students and teachers want to reduce the amount of time taken away from learning for testing and thus may appreciate unobtrusive assessment, especially when it's clear how the outputs help teachers and students immediately and directly. Yet, with regard to the second meaning, teachers, students, and parents may be rightly concerned if sensitive learning data is shared beyond the classroom without their knowledge, and that could have unforeseeable consequences for the students' performance, well-being, and future opportunities. Clearly communicating the purposes and limits of unobtrusive data collection—as well as involving teachers and students in meaningful data use—can make the difference in trust.

Due to the relational nature of trust, earning trust is also nurtured by how developers engage in communication with other ecosystem participants. Examples that developers raised in listening sessions included these:

- Trust is buoyed by transparency and disclosure.
- Trust requires effective listening and sharing.
- Trust is enhanced by demystifying why, how, and what AI is employed.
- Trust is increased when developers, educators, and researchers (among others) work together in feedback loops to identify problems and address them.
- Trust is strengthened when developers are active in forums that advance public interests.

[9] Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), Computer games and instruction (pp. 503-524). Charlotte, NC: Information Age Publishers.

Some edtech developers have already released voluntary commitments or voluntary disclosures about how they mitigate risks of AI as they develop and improve their products (see Box F).

The Department also notes that developers are already participating in public forums and initiatives where work on safe, secure, and trustworthy AI is occurring through mutual contributions of developers, educators, researchers, policymakers, funders, and more. Box D includes a list of nonprofits that are active in creating such forums. These forums are places where developers can contribute to the public interests by participating in shaping further guidelines and guardrails.

Indeed, the Department views voluntary commitments, voluntary disclosures, and participation in forums as a first step toward articulating a "dual stack" approach where developers have equally strong coordinated systems to ensure responsibility and to achieve innovation. Transparency starts with commitment but should go beyond commitment to also discussing how edtech organizations are orchestrating their development processes to be responsible from product conception to delivery and from foundational models to educational applications.

Transparency includes participating in forums but should go beyond sharing information to working together over time to build trust. The Department encourages developers to act on a belief that a healthy ecosystem requires them to step up not only with competitive innovations but also with contributions that advance the broad public interest in safe applications of AI in education.

Further, developing trustworthy edtech will likely require that developers' teams have training in AI ethics, equity, and related concerns. The Association of Computing Machinery is active in developing ethical principles for AI and other emerging technologies, and their code of ethics is designed specifically to shape the work of software developers. Trust is fortified when developers are aware of ethical principles and publicly discuss how they apply the principles in their work.

Regarding transparency, for example, LLMs are black boxes to developers and users. Further, competitive factors can overpower desires and the best intentions to communicate transparently, for example, about sources for training data. And yet, listening session attendees encouraged the Department to call for clarity of expectations: developers and educators mutually require an articulate sense of what improvements in teaching and learning can be expected. Here are some examples:

- Developers can provide opportunities to interact with their product or service and explore it deeply before an educator purchases or extensively uses it.
- Developers can provide strong training and professional development around the roles and responsibilities necessary for safe and effective use of their product or service.
- Developers can provide service guarantees around how they are available to quickly respond to and mitigate any issues that arise, including opportunities to override aspects of the product or service that are not meeting expectations.

Clarity of expectations, coupled with strong feedback loops among developers and educators, can lead to not only achieving those expectations but also exceeding them. For example, in one high-quality practice that already exists in the edtech industry, customer success managers support educators' usage with fidelity in the field, trust is fortified, and the solution provider builds a robust feedback loop for informing enhancements and updates to the solution—hopefully with fewer iterations than if designers and developers tried to innovate in a vacuum. And thus, conversations with developers in listening sessions led to an understanding of how progress through transparency is possible, highlighted in Box H.

Box H: A cycle of shared responsibility and transparency that can earn trust.

Conversations with developers led to articulating this repeatable cycle of steps that can increase trust:

- Shared responsibility begins with aligning expectations about what AI can deliver in an educational setting to be described responsibly to achieve shared educational visions, without hype or neglect of potential risks.
- Further transparency can emerge among developers and educators as they collaborate closely to address ethics, evaluate evidence, protect civil rights, address equity, and manage specific risks related to their innovations.
- Shared responsibility also increases when developers engage in forums that focus on the public interest in safe AI and where the developers can share norms, standards, and values that, if widely adopted, could make AI safer for all ecosystem participants.
- Exceeding expectations both in the AI-driven product experiences and in public contributions around responsible AI develops strong relationships with customers.
- With strong relationships around meaningful educational value and shared risk management comes trust.

Pursuing this virtuous cycle makes sense, and yet capacity may be limited by the AI literacy levels in non-developer populations. Presently, levels of AI literacy vary widely among people in the education sector, and developers can also play an important and positive role by strengthening AI literacy. Key areas for helping educators include explaining basic data governance concepts and skills, as well as how they relate to AI-specific terminology and functionality. Developers who demystify their usage of AI and provide users with an accessible concept of how AI works in their educational system show respect for their users and contribute to the growth of human agency as people adopt AI tools in education, which in turn, develops trust in the solution provider. Many learning organizations have entered the space to focus on preparing the field, and developers can learn from as well as support these initiatives for mutual benefit.

Questions To Ask

1. How can we engage with customers with balanced attention to our innovations and our responsibilities?
2. How could the NIST AI Risk Management Framework help our team to develop trustworthy AI systems?
3. What steps toward transparency could we take regarding our use of AI in products and services?
4. How can our organization contribute to AI literacy in the broader edtech ecosystem?
5. What is our long-term plan to achieve respect for our responsible use of AI in support of the overall value of our products and services to students, teachers, and others in education?

Directions to Pursue

- Developers should work to promote transparency. Developers can demonstrate a commitment can be addressed in marketing by highlighting the dual stacks of responsibility and innovation.
- Developers can share more openly written commitments and disclosures but should also emphasize two-way communication and collaborations with educators during product development and improvement.
- Developers should support efforts to build AI literacy in the ecosystem.
- Developers may consider how to publicly describe the characteristics of trustworthy system architecture achieved in their products and services. Some characteristics such as explainability may be hard to achieve in the short term, but related concepts like interpretability may be possible now while research and development continues.

Resources

- Trustworthy artificial intelligence (AI) in education: Promises and challenges | OECD Education Working Papers | OECD iLibrary
- Common Sense Media's AI Initiative
- Trust The Process: How To Choose and Use EdTech That Actually Works - EdTech Evidence Exchange
- Building trust in EdTech: Lessons from FinTech
- The Association for Computing Machinery's Code of Ethics

Conclusion

Educational decision makers express cautious optimism for new products and services that leverage new capabilities of AI. As indicated throughout this guide, educators see a wealth of opportunities to use AI to achieve the vision of their educational institutions—and yet they must be well informed of risks that must be addressed. Educational decision makers thus stress the duality of focusing on important opportunities and taking strong, clear steps to address risks. This duality shapes the market opportunity for today's educational developers.

> *"Would I buy a generative AI product? Yes! But there's none I am ready to adopt today because of unresolved issues of equity of access, data privacy, bias in the models, security, safety, and a lack of a clear research base and evidence of efficacy."*
>
> *—Patrick Gittisriboongul, ED.D, Asst. Superintendent of Lynwood Unified School District, California*

By organizing the many areas of opportunity and concern into five topics, the Department aims to sharpen developers' attention to topics of enduring importance:

1. Designing for Education
2. Providing Evidence for Rationale as well as Impacts
3. Advancing Equity and Protecting Civil Rights
4. Ensuring Safety and Security
5. Promoting Transparency and Earning Trust

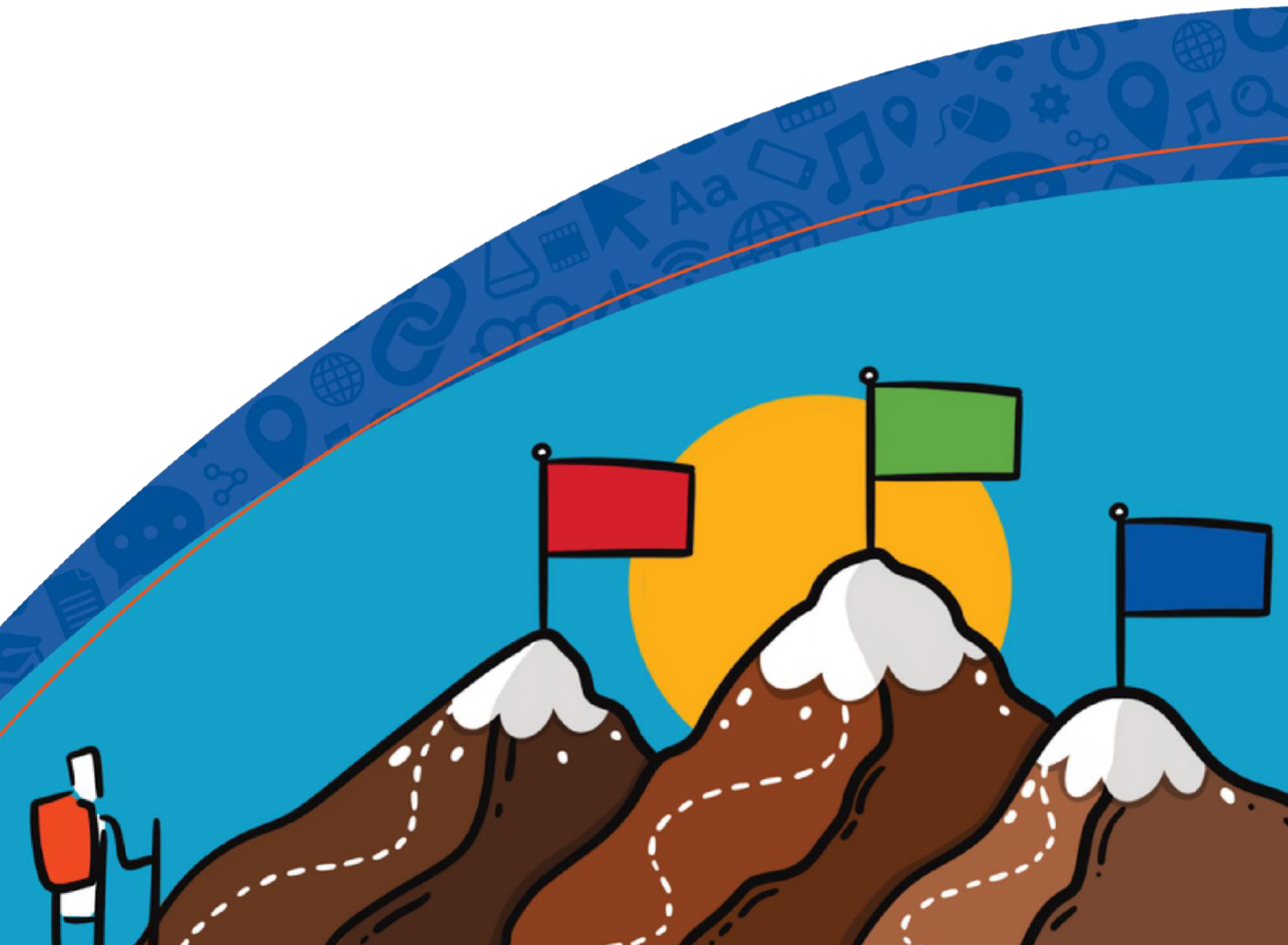
"Earning the public trust" is of paramount importance as developers launch new applications of AI in education. The Department envisions a healthy edtech ecosystem highlighting mutual trust amongst those who offer, those who evaluate or recommend, and those who procure and use technology in educational settings. The Department has found an e-bike analogy to be a good starting point for discussions across the ecosystem, offered in the Department's 2023 AI Report: that teachers and students should be in control as they use the capabilities of AI to strengthen teaching and learning. Just as a cyclist controls direction and pace but preserves energy with the assistance of an e-bike's drivetrain, so should participants in education remain in control and able to apply saved energy and time for the most impactful interactions and activities when technology amplifies their choices and actions. Now, continuing the analogy, developers should take precautions to design AI-enabled educational systems for safety, security, and to earn the public's trust, just as riders would expect e-bike developers to ensure their rider's safety and security, and to earn the public's trust.

OFFICE OF
Educational Technology

Empowering Education Leaders:

A Toolkit for Safe, Ethical, and Equitable AI Integration

October 2024



Empowering Education Leaders: A Toolkit for Safe, Ethical, and Equitable AI Integration

Miguel A. Cardona, Ed.D.

Secretary, U.S. Department of Education

Roberto J. Rodriguez

Assistant Secretary, Office of Planning, Evaluation and Policy Development

October 2024

Examples Are Not Endorsements

This document provides non-regulatory, education-specific guidance that is not intended to be comprehensive nor exhaustive. The examples and resource materials provided are intended for the user's convenience. The inclusion of any specific material is not intended to endorse any views expressed or products or services offered. These materials may contain the views and recommendations of various subject matter experts as well as contact addresses, websites, and hypertext links to information created and maintained by other public and private organizations. The opinions expressed in any of these materials do not necessarily reflect the positions or policies of the U.S. Department of Education. The U.S. Department of Education does not control or guarantee the accuracy, relevance, timeliness, or completeness of any information from other sources that are included in these materials. Other than statutory and regulatory requirements included in the document, the contents of this guide do not have the force and effect of law and are not meant to bind the public.

Contracts and Procurement

This document is not intended to provide legal advice or approval of any potential Federal contractor's business decision or strategy in relation to any current or future Federal procurement and/or contract. Further, this document is not an invitation for bid, request for proposal, or other solicitation.

Licensing and Availability

This guide is in the public domain and available on a U.S. Department of Education website at https://tech.ed.gov.

Requests for alternate format documents such as Braille or large print should be submitted to the Alternate Format Center by calling 1-202-260-0852 or by contacting the 504 coordinator via email at om_eeos@ed.gov.

Notice to Persons with Limited English Proficiency

If you have difficulty understanding English, you may request language assistance services for U.S. Department of Education information that is available to the public. These language assistance services are available free of charge. If you need more information about interpretation or translation services, please call 1-800-USA-LEARN (1-800-872-5327) (TTY: 1-800-437-0833); email us at Ed.Language.Assistance@ed.gov; or write to U.S. Department of Education, Information Resource Center, LBJ Education Building, 400 Maryland Ave. SW, Washington, DC 20202.

How to Cite

While permission to reprint this publication is not necessary, the suggested citation is as follows:

U.S. Department of Education, Office of Educational Technology, *Empowering Education Leaders: A Toolkit for Safe, Ethical, and Equitable AI Integration*, Washington, D.C., 2024.

Acknowledgements

Project Team

Empowering Educational Leaders: A Toolkit for Safe, Ethical, and Equitable AI Integration was developed under the leadership and guidance of **Roberto J. Rodríguez**, Assistant Secretary for the Office of Planning, Evaluation and Policy Development, **Anil Hurkadli**, Interim Deputy Director for the Office of Educational Technology, and **Bernadette Adams**, Senior Policy Advisor for the Office of Educational Technology.

This work was developed with support from Digital Promise under contract (91990019C0076) and led by **Pati Ruiz** and **Jeremy Roschelle** with: **Sana Karim, Babe Liberman, Gabrielle Lue, Eric Nentrup, Anthony Baker, Teresa Solorzano, Sarah Martin, and Carolina Belloc**. Graphics were developed by **Manuel Herrera** with additional figures developed by **Laura Cox**.

Contributing Members of the Educational Leader Community

- Erik Burmeister, Fremont Unified School District
- Dr. Kip Glazer, Mountain View-Los Altos Union School District
- Dr. Nneka McGee, San Benito Consolidated Independent School District
- Gina Fugate, Maryland School for the Blind
- Krystal Chatman, Jackson Public School District
- Melissa Moore, El Segundo Unified School District
- Dr. Mary Catherine Reljac, Fox Chapel Area School District
- Michael Hower, Fox Chapel Area School District
- Sallie Holloway, Gwinnett County Public Schools
- Andrew Fenstermaker, Iowa City Schools
- Dr. Patrick Gittisriboongul, Lynwood Unified
- Michael Nagler, Mineola Public Schools
- Mario Andrade, Nashua School District
- Dr. Mark Buckner, Oak Ridge School District
- Dr. Nathan L. Fisher, Roselle Public Schools
- John Malloy, San Ramon Valley Unified

Digital Promise Support Staff

- Amanda Armstrong
- Merijke Coenraad
- Diane Doersch
- Judi Fusco
- Marci Giang
- Keun-Woo Lee
- Kelly Mills
- Sierra Noakes
- Yenda Prado
- Heather Singmaster
- D'Andre Weaver
- Josh Weisgrau

Advisors

- James Basham
- Sarah Burriss
- Kristina Ishmael
- Nicole Hutchins
- Rohit Kataria
- Eleazar (Tre) Vasqez

We also thank the many educators and nonprofit organizations that attended our listening sessions and contributed their ideas for translating the Department's recommendations for Artificial Intelligence in education into practical guidelines.

Table of Contents

Introduction

The U.S. Department of Education (Department) is committed to supporting innovative advances in educational technology (edtech) to improve teaching and learning across the nation's education systems and to support educators as they incorporate emerging technology into their learning communities. Artificial Intelligence (AI) in education is a complex and rapidly expanding topic because individuals have different knowledge, opinions, and perspectives on using AI, which has implications for all members of school communities.

Who are educators and educational leaders?

In this toolkit, we use "educators" to represent all educators. This includes teachers (regardless of certification area), teacher leaders, school principals, paraprofessionals, librarians, school-based mental health professionals, and other school-based educational personnel. Representation of all educators and teaching roles in a district best supports the needs of diverse student populations.

With this representation in mind, the toolkit is meant for a broad swath of educational leaders including superintendents and school administrators, curriculum and technology leaders, school principals, educators, parents and caregivers, and engaged community members. These educational leaders can be found in a variety of learning communities.

What is a learning community?

We use "learning communities" to refer broadly to schools, school districts, and other formal learning organizations, as well as informal learning organizations such as community centers and afterschool programs. The guidance in this document is relevant for multiple settings in which students learn.

Building on the Department's prior report, Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations ("AI Report"), this toolkit is designed to help educational leaders make critical decisions about incorporating AI applications into student learning and the instructional core. The AI Report provides definitions and illustrates broad classes of opportunities, along with explaining key tensions and risks of including AI. This document sets forth and expands upon the material in the AI Report and connects broad ideas about AI to the establishment of school and district use policies that will guide its effective implementation. This toolkit provides guidance for the effective use and integration of AI in teaching and learning and presents an overview of Federal laws and considerations that are essential to anchoring and ensuring the use of AI in a safe, secure, and non-discriminatory manner. Finally, the toolkit promotes the principles of transparency and awareness in the use of AI in schools, and emphasizes the importance of providing students, teachers, and parents opportunities to opt out of AI-enabled applications in school.

Defining AI

In this toolkit, the term "artificial intelligence" or "AI" has the meaning set forth in 15 U.S.C. 9401(3): a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine- and human-based inputs to perceive real and virtual environments, abstract such perceptions into models through analysis in an automated manner, and use model inference to formulate options for information or action.

Defining Generative AI

We also use "Generative AI" (GenAI) in this toolkit. We define this term using the language from The White House Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence:

"The term 'GenAI' means the class of AI models that emulate the structure and characteristics of input data in order to generate derived synthetic content. This can include images, videos, audio, text, and other digital content."

Through listening sessions and attendance at many public events in the past year, the Department heard educators say that AI is here to stay, AI will keep changing, and safely integrating AI in educational settings will require informed leadership at multiple levels across the education system. In particular, this toolkit is informed by public listening sessions held from December 2023 to March 2024 that engaged 90 educators, who actively hold positions in PK-12 schools across the United States (e.g., classroom teachers, instructional coaches, or school

administrators), in broad discussions of the issues. Between December 2023 and March 2024, the Department's Office of Educational Technology also hosted 12 roundtable discussions on the use of AI in education with educators and educational leaders from across the nation to better understand their current perspectives and needs. Where this toolkit refers to listening sessions, it includes all these opportunities to hear from constituents.

This toolkit is organized into three distinct sections, each containing modules that can be accessed and revisited in any order depending on an educational leader's unique needs and priorities:

1. Mitigating Risk: Safeguarding Student Privacy, Security, and Non-Discrimination **(Modules 1-3).** Awareness of applicable Federal laws, rules, and regulations is an essential first step when planning for the use of AI in schools and classrooms. Educational leaders should know how existing Federal policies apply to the use of AI in their specific situations. This section invites leaders to learn about privacy and data security requirements; how civil rights, accessibility, and digital equity relate to AI; and a close consideration of the opportunities and risks associated with the use of AI. This section is relevant for an educational leader who wants to understand how proactively addressing student safety, privacy, and security can help shape their plans to use AI
2. Building a Strategy for AI Integration in the Instructional Core **(Modules 4-7).** New forms of AI have already permeated educational settings widely, and exploring AI firsthand is necessary to understanding it. Educators in our listening sessions strongly recommended that districts use the knowledge they have gained from past advances in edtech to build a clear and coherent strategy tied to the instructional core as a first step in planning for the use of AI, and then revising that strategy as they learn more about AI. That strategy should be informed by multiple sources of evidence on the use of AI. Leaders identified three additional steps for further informing their strategy for the effective use of AI-enabled tools in a manner that suits the needs of their students: (1) listen to and inform their communities, (2) establish priorities and pace for their community, and (3) guide and support implementation of a community's strategy via task force. This section provides resources to support educational leaders in considering the evidence supporting AI-enabled tools, and guiding leaders through each of these three essential steps. This path makes sense for an educational leader engaged in or beginning the strategic planning process around the use of AI.
3. Maximizing Opportunity: Guiding the Effective Use and Evaluation of AI **(Modules 8-10).** Although exploration and building coherent strategy are important early steps, the Department also urges educational leaders to be active in guiding the effective use of AI to enhance teaching and student learning, whether such tools are used for educator productivity or instruction. Educational leaders stressed three initial steps for shaping AI use: (1) developing AI literacy for educators, (2) revising responsible use policies, and (3) building a system-wide plan. This section is appropriate for an educational leader who has a clear strategy in place for the use of AI, and who is ready to focus on guiding, shaping, and continually evaluating the use of AI in their community.

Consider the metaphor of a mountain trek to represent the journey of incorporating AI in education. Like preparing for a challenging climb, achieving AI success requires careful planning, teamwork, and risk management. The trek-themed graphics in the toolkit highlight this proactive approach, reminding educational leaders of the importance of safety, ethics and equity no matter where they are on their AI journeys.

Regardless of which path an educational leader initially takes in this AI journey, we recommend navigating to the other modules in due course because the knowledge, questions, and actions in each of these three sections are designed to reinforce the others, together supporting the effective use of AI in education.[1]

1 This guide is responsive to President Biden's October 30, 2023, Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, which states:

To help ensure the responsible development and deployment of AI in the education sector, the Secretary of Education shall, within 365 days of the date of this order, develop resources, policies, and guidance regarding AI. These resources shall address safe, responsible, and non-discriminatory uses of AI in education, including the impact AI systems have on vulnerable and underserved communities, and shall be developed in consultation with stakeholders as appropriate.

Mitigating Risk: Safeguarding Student Privacy, Security, and Non-Discrimination

The modules in this section will help educational leaders weigh the opportunities and risks associated with AI-enabled tools in schools and enhance understanding of Federal laws, rules, and regulations that safeguard the rights and well-being of students as they use AI in a school setting. The modules end with discussion points and questions to facilitate conversations tailored to the local education agency. Later modules build on this content with an eye toward action. Modules 1-3 cover the following topics:

- Module 1 highlights the range of opportunities and risks.
- Module 2 reviews privacy and data security in existing Federal policy.
- Module 3 covers student civil rights, accessibility, and digital equity.

An educational leader reading this section can ask: How might these safeguards inform and shape my strategy for using AI or AI-enabled tools in my educational setting?

Module 1: Opportunities and Risks

In its AI Report, the Department notes that educational decision-making similarly requires evaluation of both opportunities and risks.

Opportunities

The AI Report highlights numerous opportunities for the effective use of AI-enabled tools to enhance student assessment, address learner variability (i.e., that all learners vary along myriad factors across the whole learner, and that these strengths and challenges can vary by context), or enhance adaptivity of academic content (personalization, differentiation, or individualization). Specific opportunities that applications of AI might also address:

- Adapting instructional approaches and content based on students' strengths and not just their needs.
- Adapting to the needs of groups who learn together and supporting social learning.
- Adapting for students with disabilities, including support for neurodiverse learners.
- Adapting to support learners on active, open, and creative tasks (whereas prior products tended to adapt for short-answer or multiple-choice tasks).
- Adapting to support self-regulation, collaboration, and communication skills in addition to more discrete academic skills.

For educators, the AI Report emphasizes how AI-enabled tools can be used to reduce administrative burden, extend instruction in accordance with a teacher's plan, or make teacher professional learning more fruitful. The AI Report also explores the opportunity for AI-enabled tools to improve feedback loops that inform students and educators about how they can improve their knowledge and skills. Opportunities include using AI to capture better information about what students know and can do, to assist in analyzing that information, and to present the information to students and educators in more useful forms.

In its listening sessions, the Department heard about many opportunities educational leaders are considering to incorporate AI-enabled tools into a strategy for student and teacher supports, including the use of AI image generators to practice understanding concepts; AI-enhanced systems supports that simplify and facilitate strategic systems planning; and AI-enabled tools that assist with lesson planning or improve classroom materials, including language translations. Several states have produced their own policy guidance about the use of AI-enabled tools in education, as have many localities. State guidance reveals additional opportunities for the effective use of AI to translate materials to familiar languages for students and their families, as well as more detail on the administrative labor that could be reduced through use of GenAI (Rochelle et al., 2024). Although it is too early to determine which opportunities will be most fruitful, the AI Report advises educational leaders to focus on using AI as an educational tool to supplement existing structures rather than replacing the role of educators and other traditional educational systems.

The AI Report also advises districts to focus on aligning uses of AI to a shared vision for teaching and learning. The Department heard from many educational leaders who affirmed the importance of a human-centered, strategic vision for education.

Risks

Balancing the multiple potential benefits of AI in education, the AI Report also presented potential risks that are worthy of consideration in the use of AI related to safety, privacy, accuracy of information, fairness, and lack of evidence. Specific tensions were noted as well, such as:

- Between collecting information to personalize vs. enabling inappropriate levels or kinds of surveillance.

- Between a teacher exercising active judgment and decision-making in the use of an AI-enabled tool vs. technology automating educational or instructional decisions.
- Between sharing information for transparency vs. protecting student privacy.
- Between attending to how context varies across education settings vs. scalable solutions for all learners.

Certain uses of AI may pose a higher risk to the rights or safety of individuals when used in educational settings. Below is a non-exhaustive list of examples where the use of AI has posed an enhanced risk to the safety, privacy, or rights of students. In these and other instances, educational leaders should carefully consider whether AI is the right solution. Because AI-enabled systems may offer to automate processes for longer periods of time or with broader scope than in previous generations of edtech, the negative consequences of unmanaged risks with AI-enabled tools can be greater. It is essential that educational leaders employ robust AI risk management practices with any AI-enhanced tool, product, or strategy.

- Detecting student cheating or plagiarism
- Influencing admissions processes
- Monitoring students' physical movements or online activity
- Projecting student progress or outcomes
- Recommending disciplinary interventions
- Determining access to educational resources or programs
- Determining eligibility for student aid or federal education or facilitating surveillance (whether online or in person)
- Utilizing biometric identification, such as facial recognition, to gate access to accounts or assignments, to take attendance, or enhance security

Risks can occur at multiple levels in an educational setting. Some risks are harmful to individual students or educators and may also constitute a violation of laws governing privacy, security, or civil rights. An individual student experiencing discrimination due to a biased algorithm is an example. Other uses and applications of AI may pose risks to the educational system or to norms important in the school culture, such as undermining the role of educators in establishing relationships with students or machine-automated instructional decisions absent important teacher judgment and humans-in-the-loop. Additional risks could undermine trust in the broader learning community. For example, trust and confidence can be undermined if a school or district uses AI-enabled tools or systems that provide inaccurate or biased information to parents and caregivers.

Educational leaders should strengthen their approaches to risk management related to the use of technology in educational settings. One starting point for strengthening risk management comes from The National Institute of Standards and Technology (NIST). This starting point is not specific to education; rather, it addresses the use of AI across industries and applications. To guide discussions about risk management, NIST created the NIST AI Risk Management Framework, which has four elements: Govern, Map, Measure, and Manage.

Govern: Develop a district culture that involves everyone in mitigating risks, enabling use of AI to advance district strategies

Map: Identify both how AI can serve important needs and lead to specific risks by key factors such as grade-level, subject matter, and knowledge of the students and the setting.

Measure: Prioritize risks, and work with AI product vendors and with locally-gathered evidence to measure both the positive impacts and degree of risk mitigation.

Manage: Strengthen capacity to manage implementation of AI tools to realize desired benefits while managing both anticipated and emergent risks.

Module 10 of this toolkit suggests how to use these risk management elements in an educational setting.

The following non-exhaustive list provides real-world examples of various risks associated with the use of AI in schools and educational settings. As AI technologies continue to evolve and as their use expands, additional and unanticipated risks may also emerge.

Transparency About AI Risks: In some cases, educational leaders have not been able to obtain transparent information about how their AI-enabled too ls or systems function and collect data including essential information about the tool or system's performance, evaluation practices, and processes to mitigate bias. This lack of information presents a significant challenge in providing transparency to educators, parents, students, and their wider educational communities.

Bias and Fairness: Some test proctoring systems that rely on facial recognition algorithms have unfairly and disproportionately flagged non-white students and students with disabilities for cheating. AI in education has shown potential to reinforce existing inequities, biases, and unfairness due to biased training data, algorithmic discrimination, and insufficient diversity in AI development. Product developers have more recently undertaken efforts designed to address these issues, including improving data diversity, developing fairness-aware algorithms, and promoting diversity in AI development.

Harmful Content: Students often use AI to create images for their school projects. However, GenAI has been found to perpetuate negative and harmful stereotypes based on race, sex, and disability when given prompts to construct images representing different student cohorts.

Malicious Use: GenAI has been used to perpetuate cyberbullying by students, including creating false and negative narratives about others, or generating "deep fake" fabricated images depicting peers and later posting these images on social media. Some have also used AI to impersonate officials in communicating malicious, false, or harmful messages to the broader school community.

Hallucination Risk and Wrong Information: Despite ongoing research, GenAI has still been found to produce inaccurate or factually incorrect outputs at times, such as narratives describing historical figures who never existed or wrong answers to math problems.

Overreliance: Educators who rely on AI-enabled tools and systems to identify students who may be "at risk" and in need of extra support might sometimes overlook valuable insights regarding context, background, or alternate risk factors not recognized or accounted for in these systems.

Urgency to Adopt: GenAI chatbots became widely available to schools before adequate guidance was provided to help educators and students use them responsibly. This distracted educators from managing the core functions of teaching and learning. Educational leaders have faced pressure to allow various AI-enabled technologies into schools due to rapid interest and the fear of falling behind.

Managing Risks

Additionally, we provide a non-exhaustive list of practices to manage the risks of AI use to protect rights and safety in the use of AI. These practices mirror the minimum risk management practices that Federal agencies must implement in their own use of rights- or safety-impacting AI. For example, the Office of Management and Budget recently put out Memorandum OMB M-24-10, "Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence," which directs agencies to manage AI risks and ensure that AI applications are safely deployed, transparent, and aligned with ethical standards. We recommend similar considerations for educational leaders to ensure responsible AI use in education and to safeguard student rights and privacy. As AI technologies evolve, so too should effective practices for mitigating risk in the use of AI.

1. **Complete an AI impact assessment**. Educational leaders should document the intended purpose for the AI and its expected benefit; the potential risks of using AI; and the quality and purpose of the data used to develop, train, test, and operate the AI.
    a. Especially for predictive AI models, the decision about which intervention to use based on the AI model is a major factor in ensuring effective and equitable outcomes. Models based on historical data are insufficiently reliable and carry greater risk for impacting rights if used for punitive interventions (such as disciplinary action) or significant opportunities (such as admissions decisions) as opposed to supportive interventions (such as tutoring services).
2. **Test the AI for performance in a real-world context**. Educational leaders should conduct tests to demonstrate that the AI will achieve its expected benefits and that associated risks will be sufficiently mitigated. Testing conditions should mirror as closely as possible the conditions in which the AI will be deployed.
    a. Educational leaders are also encouraged to leverage pilots and limited releases with strong monitoring, evaluation, and safeguards in place to carry out the final stages of testing before a wider release.

3. **Independently evaluate the AI**. Educational leaders should obtain independent assessments that the system works appropriately as intended and that its expected benefits outweigh the potential risks.
4. **Conduct ongoing monitoring**. In addition to pre-deployment testing, educational leaders should institute ongoing procedures to monitor degradation of the AI's functionality and to detect changes in the AI's impact on rights or safety.
5. **Regularly evaluate risks from the use of AI**. The ongoing monitoring process should include periodic human reviews to determine whether the deployment context, risks, benefits, and agency needs have evolved. Human review is especially helpful after significant modifications to the AI or to the conditions or context in which the AI is used.
6. **Mitigate emerging risks to rights and safety**. Educational leaders should take steps to mitigate risks identified through ongoing monitoring or periodic review. Steps may include updating the AI to reduce its risks or implementing procedural or manual mitigations, such as more stringent human intervention requirements.
7. **Ensure adequate human training and assessment**. Educational leaders should ensure there is sufficient training, assessment, and oversight for operators of the AI to interpret and act on the AI's output, combat any automation bias, and ensure human-based components of the system effectively manage risks from the use of AI.
8. **Provide additional human oversight, intervention, and accountability as part of decisions or actions that could result in a significant impact on rights or safety**. Educational leaders should determine whether there are decisions or actions in which the AI is not permitted to act without additional oversight and implement appropriate human oversight, intervention, and accountability accordingly.
9. **Provide public notice and plain-language documentation**. Educational leaders should provide reasonable and timely notice to affected populations when an AI system is being used. Educational leaders should also provide information to members of the educational community on how the AI was trained and tested, and on how risk mitigation practices were implemented.
10. **Identify and assess AI's impact on equity and fairness and mitigate algorithmic discrimination when it is present**. Educational leaders should determine whether the AI model results in significant disparities in the model's performance across demographic groups. Educational leaders should mitigate disparities that lead to, or perpetuate, unlawful discrimination or harmful bias.
11. **Consult and incorporate feedback from affected communities and the public**. Educational leaders should consult affected communities, including underserved communities, and solicit public feedback in the design, development, and use of AI.
12. **Conduct ongoing monitoring and mitigation for AI-enabled discrimination**. Educational leaders should monitor for AI-enabled discrimination against protected classes during ongoing monitoring.
13. **Notify negatively impacted individuals**. Educational leaders should notify individuals in a clear and accessible manner when the use of AI results in an adverse decision or action that specifically concerns them, such as placement in a particular class or identifying cheating within a student's response to an assignment.

14. **Maintain human consideration and remedy processes**. Educational leaders should provide a fallback and escalation system in the event that an impacted individual would like to appeal or contest the AI's negative impacts on them. These remedy processes should not place unnecessary burden on the impacted individual.
15. **Maintain options to opt out for AI-enabled decisions**. Educational leaders should provide and maintain a mechanism for individuals to conveniently opt out from AI functionality in favor of a human alternative. Opt-out mechanisms should be prominent, readily available, and accessible.

Discussion Points and Questions

These discussion points and questions are intended to catalyze conversations within learning communities regarding Opportunities and Risks.

Discussion Points:

- Discuss why it is important to frame the conversation about using AI in terms of both opportunities and risks.
- Invite members of your learning community to share opportunities they have explored and are excited about.
- Involve members of your learning community in elaborating their concerns about risk. Use examples to illustrate that risks can extend beyond privacy and data security and include significant negative impacts to individual people and the integrity and trust of the learning community.
- Explore how your learning community presently manages risks due to use of technologies and how this approach could be upgraded to address the breadth and depth of risks that may emerge as AI is adopted.

Discussion Questions:

- Which opportunities to use AI best align to values and vision for the educational system that our educational leadership team serves?
- How can our educational leaders obtain information about risks related to each educational opportunity, for example, from students, educators, community members, vendors, researchers, or through local risk study teams?
- How can the recommended minimum risk management practices and the NIST AI Risk Management Framework inform the processes our leadership team will establish locally to manage risk?
- How can our leadership team manage the tensions between desire among students, parents, or staff to use (or not use) AI and incomplete or worrisome information about risks?
- How can we make sure that vulnerable populations are not disproportionately affected by the risks of AI?

Module 2: Privacy and Data Security

In order to ensure the effective use of AI in education, educational leaders must be well versed and current in their knowledge regarding relevant laws, rules, and regulations related to privacy and data security. In order to maintain compliance and safeguard their communities of students, families, educators, and school staff, the Department regularly provides information about privacy and data security legal requirements and best practices through its Student Privacy Policy Office (SPPO). SPPO offers extensive information on protecting student privacy, including protecting students' education records and the personally identifiable information (PII) contained therein.

> Visit the SPPO website for resources offered through SPPO's Privacy Technical Assistance Center (PTAC) that address the Family Educational Rights and Privacy Act (FERPA) and Protection of Pupil Rights Amendment (PPRA).
>
> Visit the Department's Office of Educational Technology website for content and events to support edtech developers in safeguarding student data.

In addition to updating their knowledge and understanding of state and local privacy and data security laws, educational leaders can begin their work on privacy and data security when using AI by reviewing and updating their understanding of the following Federal laws and policies:

- Children's Online Privacy Protection Act **(COPPA)**: Enforced by the Federal Trade Commission (FTC), COPPA imposes requirements on operators of websites or online services directed to children under 13 years of age, and on operators of other websites or online services that knowingly collect personal information online from a child under 13 years of age. Presently, FTC is engaged in new rulemaking which considers how industry developers incorporate children's data in AI technology creation and in the deployment of AI systems, who should be authorized to access and use that data, and how these systems interact with children.
- Family Educational Rights and Privacy Act **(FERPA):** Educators must consider whether the information about students shared with or stored in an AI-enabled system (i.e., information about student background or interests) is subject to Federal privacy laws such as FERPA. Enforced by the U.S. Department of Education, FERPA governs the privacy of students' education records and the PII contained therein, as maintained by educational agencies and institutions (such as school districts, colleges, and universities) or by parties acting for such agencies or institutions. FERPA applies to agencies and institutions that receive funds under any program administered by the U.S. Department of Education. Among other things, FERPA generally requires educational agencies and institutions to obtain prior written parent or "eligible student" consent before disclosing such records and PII, with certain defined exceptions. Certain edtech products may utilize AI or other technologies to collect lengthy back-and-forth dialogues with

students; absent the proper safeguards, these exchanges and data may contain more sensitive information collected without student or parent consent.

- Individuals with Disabilities Education Act **(IDEA):** Implemented by the U.S. Department of Education, this law enacted in 1975 governs special education and related services provided to children with disabilities, as well as early intervention services provided to infants and toddlers with disabilities and their families. IDEA contains confidentiality requirements to protect the privacy interests of children with disabilities from birth until age 21 who are referred for services. Like FERPA, IDEA requires that a parent provide prior written consent before PII is disclosed to a third party and that educators obtain informed consent from parents (with some specific exceptions). For example, "alt text" can support understanding information from images, and with GenAI, it will become possible to synthesize or customize alt text to the needs of specific learners; however, issues related to the unauthorized release of PII through use of alt text, such as data being shared with third-party vendors, will need to be addressed. The toolkit discusses IDEA further in the "Civil Rights, Accessibility, and Digital Equity" module.
- Children's Internet Protection Act **(CIPA):** The Federal Communications Commission provides oversight of CIPA, a law that imposes requirements on schools or libraries that receive E-Rate discounts for internet access. CIPA requires schools receiving discounts offered by the E-Rate program to adopt an internet safety policy addressing unauthorized disclosure, use, and dissemination of personal information regarding minors. Such safety policies will likely need updating as the value of data about students increases and cybersecurity threats correspondingly increase. The Federal Communications Commission's CIPA Guide offers a more in-depth understanding of CIPA requirements.
- Protection of Pupil Rights Amendment **(PPRA):** Enforced by the U.S. Department of Education, PPRA affords parents of students certain rights regarding, among other things, participation in surveys and the collection and use of information for marketing purposes. PPRA requires local educational agencies receiving Federal funds from the Department to develop and adopt policies, in consultation with parents, that afford parents the right to inspect any instructional materials (excluding academic tests or academic assessments) used by the local educational agency as part of the educational curriculum, upon request.

As they implement a strategy to guide the use of AI in schools, educational leaders will likely need to seek further guidance on data privacy and security in order to ensure compliance with these and other statutes. Many reputable organizations, including non-governmental and nonprofit organizations, provide additional information that could be helpful.

Monitoring Privacy and Security with AI

Educational leaders have effectively addressed student privacy and security in the wake of prior waves of technological advancement and use of edtech tools in schools. However, with the advent of AI-enabled technologies, issues related to student privacy and data security can become significantly more complicated and involved. For example:

- AI-enhanced products and services may inappropriately or even inadvertently collect or generate student data that contains PII.
- AI-enhanced products and services may not adequately answer the questions parents, students, educators, and others may have about data collection and use (or where necessary, obtain consent from the relevant user).
- AI-enhanced products and services may not be transparent in disclosing how and why they collect student data or PII, and educators, students, or parents may not be provided an affirmative option to opt out of data collection.
- AI-enhanced products may not adequately protect student data, resulting in wrongful use or disclosure of student information.
- AI-enhanced products could result in students under the age of 13 accessing inappropriate or harmful content.

Further, within emerging applications that incorporate AI, these issues may intensify, and new threats may appear. Drawing on information relayed in listening sessions with educators and in public conferences, the Department notes the following enhanced concerns, separate and apart from potential violation of applicable privacy laws:

- AI-enhanced products may collect new and highly sensitive forms of data, such as a students' physical characteristics or their whereabouts, and even the questions that students informally ask. This is in contrast to more traditional edtech tools and products, which collected more limited or discrete input, such as keyboard touch and mouse input, in response to structured educational prompts. Disclosure of a student's voice or face could result in harm if, for example, synthesized content shows them falsely expressing something they did not say through methods such as "deep fakes."
- AI-enhanced products may collect extensive data about students' activities and locations, enabling new access into students' lives and broader activities. This data enhances the potential for individual surveillance and, if data are leaked, could result in bad actors' use of these details to harass, bully, or otherwise harm students.
- AI-enhanced products collect high volumes of student data, and applying algorithms to this volume of data may enable making broad inferences about students' future behavior, such as what kinds of products they might later buy. Further, even data without obvious PII may disclose a student's identity because identity can be inferred from a large volume of data points. Experts have already reported seeing an increase in scale and frequency of cyberattacks in schools due to the increasing value of student data for non-educational uses (United States Government Accountability Office, 2022).

- AI-enhanced products may collect data through communications and interfaces that mimic human social interaction, including chats and other human-like forms of interaction. This may result in educational participants being less careful about what they input. For example, a well-intended teacher could inadvertently put sensitive and private information about a student's disability into an unprotected or unencrypted chat interface in order to obtain a customized lesson plan or an Individualized Education Program (IEP), without being aware of how the data could be used by the developer or by others granted access to the data.
- In AI-enhanced products, it may be difficult to inspect how an algorithm or AI system collects and makes use of data, since data is not being collected via more traditional or familiar means such as fixed surveys, fields within the curricular pages, or other pre-determined formats. The data in these systems may be a "black box." For example, an AI-based product might synthesize a new dialogue that obtains PII from students, even if the dialogues previously observed in a product demonstration or inspection did not do so.

Discussion Points and Questions

These discussion points and questions are intended to catalyze conversations about Privacy and Data Security.

Discussion Points

- Identify relevant Federal, state, and local laws and their key provisions. Describe steps your institution has already taken with regard to Federal, state or local laws with AI or prior waves of edtech, and what you plan to do now.
- Discuss community concerns about privacy from the various roles and perspectives within the school district. This should include representation from the district's various demographic groups, educators and staff across different teaching assignments and years of tenure, and community members including students' family or guardians.
- Review past actions to address privacy and data security, such as actions taken in response to data breaches, or previous plans to address potential breaches.
- Elaborate ways in which risks are intensifying in an age of AI. Consider why bad actors may have stronger incentives to breach data privacy, the new kinds of private information that AI-based systems may store, and the quantity of detailed information about students' activities that may be available.

Discussion Questions

- With AI use increasing, what new kinds of privacy incidents might occur in our educational settings? How might guidance from trusted learning organizations be leveraged to design responses when an incident does occur? For example, some educational leaders leverage resources developed by The Consortium for School Networking (CoSN), including their tools and resources.

- For existing products widely used in our schools, how is AI being incorporated in each respective vendor's product roadmaps, and how transparently can they explain any changes to privacy and data security risks? For example, consider the sample vendor letter shared by Advanced Learning Partners.
- What key privacy and data security risks do we observe when students, educators, and community members use products that include AI? For example, how are risks, such as those in the Center for Democracy & Technology report, being experienced and addressed?
- How might we use existing communication channels to inform our community about privacy and data security risks in a way that respects their concerns and provides information and resources that build confidence about privacy and data security?
- How might we support educators and other adults to protect student privacy and security? For example, implementing resources developed by CoSN including the Student Data Privacy Toolkit.

Module 3: Civil Rights, Accessibility, and Digital Equity

Fairness is an important American value, upheld by laws and educational policies that promote a national culture of equitable systems. Educational leaders should understand interplay between at least three key areas of policy to practice fairness in their schools or districts as it relates to AI: Civil Rights, Accessibility, and Digital Equity. This module highlights Federal laws, rules, and regulations as they relate to Civil Rights, Accessibility, and Digital Equity; there is potential overlap across the three topics.

Civil Rights

The Department's Office for Civil Rights (OCR) enforces civil rights laws that protect students against discrimination based on race, color, national origin, sex, disability, and age. These laws prohibit discrimination inside and outside the classroom. OCR handles a wide variety of cases and responds to potential violations of civil rights law, including those related to discriminatory discipline, racial harassment, unequal access to educational resources, denial of language services to English learners, and other denials of equal educational opportunities based upon race, color, national origin, sex, and disability. Across all of these civil rights laws, public schools and universities receiving federal funds, as well as private schools and universities receiving federal funds, are required to address discrimination that unfairly blocks students from participating in or learning from educational activities. For example: Title VI of the Civil Rights Act of 1964 prohibits discrimination on the basis of race, color, or national origin; Title IX of the Education Amendments of 1972 prohibits discrimination based on sex; and the Age Discrimination Act of 1975 prohibits discrimination based on age in programs receiving federal funding. Together, these statutes emphasize the need for AI in education to be inclusive, free from bias, and accessible to all across various demographics.

Accessibility for Individuals with Disabilities

On April 24, 2024, the Department of Justice issued its final rule revising the Title II regulations for implementing the Americans with Disabilities Act (ADA). The rule establishes specific requirements, including the adoption of technical standards to make services, programs, and activities offered through web and mobile applications by state and local government entities accessible to people with disabilities. This includes content provided by public schools and universities. This regulation adopts an internationally recognized accessibility standard for web access, the Web Content Accessibility Guidelines ("WCAG") 2.1.

The U.S. Department of Education's Office of Special Education and Rehabilitative Services administers programs and services supporting millions of children, youth and adults with disabilities, including programs authorized under the IDEA. The IDEA assists states and public agencies in the delivery of early intervention, special education, and related services to more than 7.6 million (as of school year 2022–23) eligible infants, toddlers, children, and youth with disabilities. In addition to early intervention services to eligible children, ages birth through 3, the law makes available a free appropriate public education to eligible children with disabilities throughout the nation and ensures special education and related services to those children. The

IDEA also provides substantive protections on various aspects of a student's educational placement and experience, including the development of an IEP, the delivery of appropriate services, review of potential disciplinary actions, and more that may be impacted by the adoption of AI tools in the educational setting.

The IDEA highlights the importance of educational resources that leverage students' strengths and resources to address their needs. The Department provides guidance and technical assistance on improving results and outcomes for children and youth with disabilities served under IDEA and Section 504. This includes resources such as the Center for Innovation, Design, and Digital Learning (CIDDL), which has provided guidance on the application of AI in education, specifically in special education in their Inclusive Intelligence: The Impact of AI on Education for All Learners report. Additionally, Section 504 of the Rehabilitation Act of 1973 prohibits discrimination on the basis of disability by recipients of federal financial assistance, and Title II of the ADA prohibits discrimination on the basis of disability by public entities whether or not they receive such financial assistance. In the context of AI, IDEA and Section 504 underscore the need to design tools that are inclusive, adaptable, and compliant with accessibility standards, ensuring fair and effective support for students with disabilities. Learn more about these statutes on the U.S. Department of Education website.

Digital Equity

Digital equity ensures that everyone has fair access to technology, internet connectivity, and digital literacy resources, regardless of socioeconomic status, geography, or other factors. Recent legislation, including the Bipartisan Infrastructure Law and the Digital Equity Act, aims to close the digital divide by funding broadband access and digital skills programs across underserved communities. Achieving digital equity is essential to maximizing the benefits of AI in education, ensuring all students and educators can access and effectively use digital tools to enhance learning.

The National Educational Technology Plan (NETP), along with other related Department publications, includes guidance on addressing digital equity. NETP identifies three distinct but interconnected equity "divides" to better understand the relevant issues, all of which are discussed in more detail below.

- The **Digital Access Divide** addresses unequal access to hardware, connectivity, content, and experiences using technology in education.
- The **Digital Use Divide** addresses differential access to high-quality uses of technology that provide students with meaningful opportunities to explore, create, and engage in critical analysis of academic content and knowledge.
- The **Digital Design Divide** addresses allocations of professional development and resources to involve educators in designing appropriate uses of technology for their students.

The three digital divides outlined in the NETP express themselves in new ways with AI, as follows:

- **Digital Access and AI.** Building and running the underlying models that power GenAI are presently very expensive, and the business models for recovering these costs are not fully established. Additionally, subscription models for premium services are often affordable only to some schools and/or families and not others. Likewise, GenAI-based tutoring models that require high bandwidth access to a Large Language Model (LLM) may generate high per-student costs, which can result in unaffordable pricing for most students. Although GenAI may appear to be free, it is not, and the costs may be distributed in ways that present a barrier to equitable access.
- **Digital Use and AI.** Uses of AI may differ depending on multiple factors, including existing digital access or access to financial resources. In listening sessions and at national conferences, the Department has heard concerns from educators that students in schools serving higher-poverty communities with fewer resources will experience AI as a cost-cutting measure and substitute for human educators and mentors, while those attending schools serving wealthier communities will experience AI as augmenting what human educators provide. Alternatively, AI may be used exclusively in remedial roles with some students, while in more expansive and developmental roles with others. In addition, due to bias in AI-systems, educators are concerned that students with disabilities or students of color may have a higher chance of being accused of using AI inappropriately, such as overusing AI. Educational leaders should seek evidence to ensure that any negative consequences of using or misusing AI do not disproportionately impact students from protected groups. Leaders should seek evidence that AI use is fair to all students.
- **Digital Design and AI.** In its AI Report, the Department recommends centering educators in the design of AI-enabled tools and systems and in how AI is used in education. AI Literacy (see Module 8) is a prerequisite for building the capacity of educators to contribute in a meaningful and effective manner in the design of educational experiences with technology. Additional professional learning relevant to the specific educational content, pedagogical approach, and technology should also be provided to educators. Engaging educators in working groups or task forces is one way to accomplish this goal. In general, educators need time, access, knowledge, and dedicated resources to participate in design.

These examples point to the general fact that AI is entering learning communities that have existing digital divides impacting student opportunities and learning. For more than a decade, schools have used AI-related systems to predict student test scores, reading levels, and other achievement characteristics. School leaders have seen both the benefits and the potential bias in such systems. Local educational leaders will need to elaborate how the features of their specific system will come into play as AI becomes prominent in their local educational practices—and what they can do throughout their system to advance digital equity.

Connecting AI with Educational Equity

Accessibility and Assistive Technology

Universal Design for Learning (UDL) is a set of widely recognized evidence-based learning principles and a "framework to improve and optimize teaching and learning for all people based on scientific insights into how humans learn" (CAST, 2024). UDL-informed AI tools can support learning for all students by providing multiple means of representation, engagement, action, and expression. Federal laws such as IDEA, the Every Student Succeeds Act (ESSA), and the 21st Century Assistive Technology (AT) Act call for equitable access to assistive technologies for students with disabilities and learning differences. The Department's Office of Special Education Programs (OSEP) provides resources such as training and consultations to encourage assistive technology use that support student access, learning, and assessment. See the OSEP website to learn more.

AI can accelerate the integration and use of UDL and AT into teaching and learning to reduce learning barriers, benefitting students with disabilities and neurodiverse learners. Importantly, GenAI is evolving to support multimodality—the combination of various communication modes to create meaning and applies in many contexts. With AI, multimodality presents itself in technology interfaces that can listen to speech, provide captioning, translate languages, and even recognize American Sign Language. Further, AI may assist students with visual impairments by producing reliable "alt text" that describes images. It can render increasingly higher quality speech-to-text and can carry out dialogues with students through conversation and accelerate UDL. In turn, multimodality could power important new assistive technologies.

However, to date, accountability for accessibility and WCAG standards for educational resources is uneven. Therefore, educational leaders may need to urge developers using AI to build on their strengths and meet the needs of their learners.

Bias and Algorithmic Discrimination

With the growing use of AI in schools and its ability to operate on a mass scale, schools may create or contribute to discrimination.

Algorithmic discrimination happens when automated systems treat people unfairly, or negatively impact them, because of their race, ethnicity, gender, religion, disability, or other legally protected characteristics. AI models in education have the potential to exhibit algorithmic discrimination when used for tracking, for example, student preparedness for graduation or student discipline issues, in any predictive way based on historical data. This is because the AI systems were trained on historical data that could proliferate historic and systemic biases that have existed in our education systems and learning communities.

As deployers of automated systems, educational leaders should take proactive and continuous measures to protect students, educators, and learning communities from algorithmic discrimination, and to use systems in an equitable way.

Algorithmic discrimination highlights how biases embedded in technology can lead to unfair treatment, but these biases are not limited to machines. Educators themselves may also exhibit societal and cultural biases and discrimination when evaluating, and disciplining, students on their AI use. While concerns about AI tools, such as ChatGPT, facilitating cheating are widespread among educators, research has so far demonstrated that students do not use AI to as large an extent as teachers assume (Center for Democracy and Technology, 2023). This research has also shown that a minority of teachers have received guidance on detecting student use of tools like ChatGPT for cheating. Together with human bias, this insufficient guidance could result in disproportionate disciplinary actions against historically marginalized students and erode trust between students, families, and educators.

Below are examples of concerns related to bias and discrimination in the use of AI in education, raised by educators and expressed in the Department's listening sessions and public conferences. Similar situations could arise if educational leaders give decision-making autonomy and power to AI systems and tools without monitoring bias and algorithm discrimination.

- AI applications may monitor students and teachers based on their online activity, screen time, and physical activity when using the school internet and devices, and through surveillance cameras with facial recognition and other AI-enabled capabilities.
- An AI application that automates course selection or career recommendations may systematically deter students from pursuing courses or careers that they could succeed in based on historical biases that suggest that "certain students didn't succeed in the past."
- An AI application that detects student behavior, for example, when taking a test or in a school hallway through computer-based vision, may unfairly single out students of color for disciplinary action. An application that employs the AI subfield of computer vision to interpret images or video from a device's camera or a school's security camera may have been trained on material that is biased, for example, against students of color and students experiencing poverty, and has not been reviewed by humans in our current context. Unchecked, these biases may show up in reality, for example, by disproportionately identifying students of color as requiring disciplinary action.
- An AI application that sequences online mathematics modules for a student may direct English learner students to remedial modules more often. The English learner students may know the math but not perform well with the reading load that surrounds the math.
- An AI application could block educational progress by discriminating against students from a particular background by repeatedly failing to recognize their speech in English and asking them over and over to reiterate their input (whereas a human hearing the same speech may understand it the first time).
- An AI image generator might provide images drawn from racist, sexist, and/or ableist stereotypes when students are using the tool for creative assignments.

Discussion Points and Questions

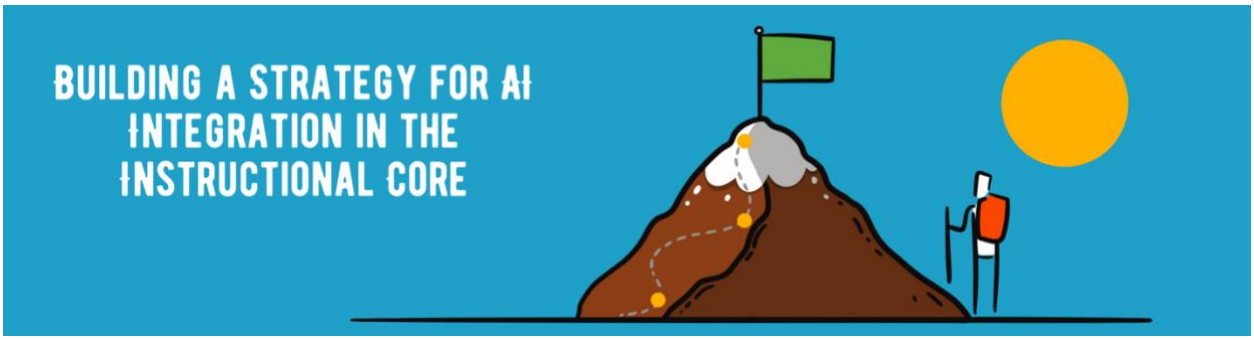
These discussion points and questions are intended to catalyze conversations among your learning communities with regards to Civil Rights, Accessibility, and Digital Equity.

Discussion Points:

- Introduce and explain key concepts such as bias, algorithmic discrimination, multimodality, and digital divides, and how these concepts can clarify the relationship between equity and AI.
- Explicitly consider the biases and inequities that AI tools—specifically those used for monitoring and surveilling student digital, online, and in-person activity—can have on historically marginalized students.
- Provide guidance for AI educators on how to build trust with their students and detect and respond to responsible and irresponsible student use of AI tools. This approach is essential for preventing bias, avoiding the disproportionate disciplining of students from historically marginalized backgrounds, and reducing the risk of distrust between students, their caregivers, and the educator.
- Understand relevant areas of Federal law and policy, e.g., civil rights, accessibility, and the digital divides, and discuss how they currently show up in your local educational setting as you adopt AI and other emerging technologies.
- Ask special education educators to offer examples about how multimodal features could be helpful to specific populations and also what risks should be mitigated for the populations they serve.
- Explore examples of the three digital divides and discuss how your institution addresses them currently and when emerging technologies are adopted.

Discussion Questions:

- What questions should district leadership ask AI developers regarding their assessment and correction of bias and algorithmic discrimination in their AI products?
- Where are algorithmic biases and discrimination risks most pronounced and potentially consequential as a district's leadership team considers using AI to automate decisions in school systems, including teaching and learning decisions as well as student support and logistics decisions, and monitoring online activity?
- How can district leadership and educators use AI in a way that is not surveilling students but prioritizes their safety?
- How can district leadership provide guidance to educators to effectively evaluate student use of AI and effectively respond to student misuse so that educators are not disproportionately disciplining their students?
- What can a district's leadership team do to address the three divides discussed in the NETP, as they apply in our educational setting?



Building a Strategy for AI Integration in the Instructional Core

The modules in this section are designed to support educational leaders in establishing a strategy that supports the effective use of AI in advancing the instructional core, providing room to address both current and future AI technologies.

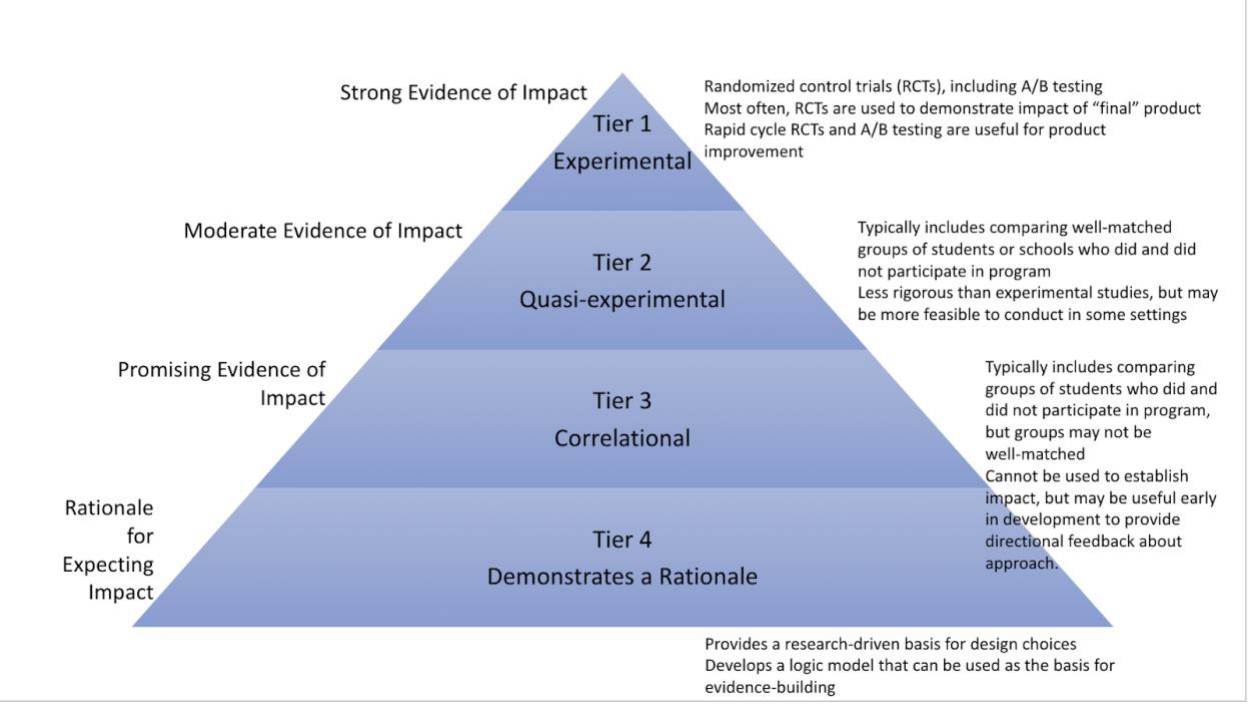
- Module 4 emphasizes the importance of **evidence**.
- Module 5 considers the **instructional core.**
- Module 6 recommends that educational leaders begin their strategic planning by developing awareness and knowledge, through **informing and listening**. Educators can counter the perceived rush to apply AI in diffuse ways, with unknown consequences, by **setting priorities and pacing** for their community.
- Module 7 urges educational leaders to establish a task force to provide responsive **guidance and support** for educators, students, parents and caregivers, and the larger learning community who are using AI in educational settings.

An educational leader who is beginning to read this section can ask: How might I leverage a range of data and perspectives to build a strategy for responsible use of AI in my educational community? How might I deepen my understanding of the evidence and the connection between AI and the instructional core? How might I set a pace for implementation and clear goals for success?

Module 4: Understanding the Evidence

In 2015, the Every Student Succeeds Act (ESSA) reauthorized the Elementary and Secondary Education Act (ESEA) of 1965. ESEA both encourages and requires school and district decision makers to choose evidence-based educational products and services ("interventions") that have been shown to improve student outcomes through high-quality research and evaluation, including research that demonstrates a rationale for the use of technology. ESEA defines four

tiers of evidence. The highest quality of evidence is referred to as "Strong Evidence," and the weakest evidence is referred to as that which "Demonstrates a Rationale." Types of evidence at each tier are described in the graphic below. In considering the topic of evidence, educational leaders should review both the kinds of evidence discussed in Federal policy (discussed below) and also evidence from community members' lived experience (Module 6).



Educational leaders should understand the characteristics of sound evidence in each tier and are encouraged to use the evidence tiers as they evaluate adoption and responsible use policies of both AI and non-AI products, tools, and services in their learning communities.

Additional federal resources provide information regarding use of evidence. The What Works Clearinghouse (WWC) reviews research on educational resources, determines which education research studies meet rigorous standards, and summarizes education research findings. For some applications of AI, such as intelligent tutoring, the WWC has existing reviews detailing key findings within educational research. Newer products, however, may not yet be specifically reviewed. WWC Practice Guides summarize research-based principles with broad applicability, and these can be very helpful for understanding principles that could be incorporated into a product's design. In addition, the Office of Educational Technology offers an Edtech Evidence Toolkit with resources including one-pagers, case studies, and examples to support educational leaders in making evidence-informed edtech adoption decisions within schools. More generally, Federal agencies including the Department, the Institute of Education Sciences, the National Science Foundation, and the National Institutes of Health fund research to evaluate and publish technological approaches to improving education.

For educational leaders who may find it difficult to comb through the extensive literature on their own, forming a research-practice partnership with a researcher may be helpful. Research-practice partnerships are collaborative relationships between researchers and practitioners that aim to address real-world challenges in education through the practical and collaborative application of research in educational settings (Coburn & Penuel, 2016; Farrell, et al., 2022). The Department's Institute of Education Sciences, the National Science Foundation, and certain philanthropic organizations (such as The Lumina Foundation, The New School Venture Fund and The Bill & Melinda Gates Foundation) all fund research-practice partnerships. Research-practice partnerships between educational leaders and researchers are typically designed to a) promote evidence-based practices in schools, b) bridge the gap between academic research and practical application in educational settings, and c) facilitate the translation of research findings into actionable strategies to positively impact learner outcomes.

Connecting Evidence to AI-Enabled Interventions

Research on AI in education has a 50-year history, providing sound evidence for some applications of AI in teaching and learning. For example, many in the field have identified AI-based tutoring as a promising application, and Intelligent Tutoring Systems (ITS) have been studied for a long time (Steenbergen-Hu & Cooper, 2013). Indeed, meta-analyses (statistical summaries of many independent studies) exist for AI-based mathematics tutoring (Kulik & Fletcher, 2016; Xu et al., 2019).

Successful interventions described in the literature might include elements that newer GenAI-based tutors lack. For example, GenAI-based tutors may not have information about how to sequence instructional content, how to diagnose common student misconceptions, or how to identify strengths in student knowledge that can enable the student to learn a target idea. In another instance, the literature reveals (VanLehn, 2011) that tutors are especially beneficial when they can intervene on a specific problem-solving step that a student got wrong, rather than intervening only when the overall answer to a problem is wrong. How to use AI and machine learning to intervene appropriately with different students is an active research issue (Gao et al., 2024) as is the more general issue of how to incorporate evidence-based pedagogies into the response strategies of GenAI-based tutors (Weiss, 2024).

Because newer GenAI tutors may not intervene at the detail level of a problem-solving step (Chine et al., 2022), prior evidence may be helpful in probing areas where the newest products may fall short of what has been shown to work.

Many applications of advanced technology are initially exciting, but when evidence is collected to evaluate the application, no or minimal benefits are detected. Educational leaders should carefully look at the quality of evidence for novel applications of AI in education. Given the rapidly changing landscape of AI-based products, evidence at any of the four tiers could be useful to support educational leaders' decision making. We provide examples below:

- **Tier 4 (Demonstrates a Rationale):** Educational practices, interventions, program components, and tools at the Tier 4 level have a well-defined logic model, are supported

by prior research with positive findings (such as the WWC Practice Guides) and have efforts underway to determine their effectiveness. Methods incorporating literature reviews, development of logic models, pilot testing, and surveying/user feedback are all examples of developing a Tier 4 evidence base.

- **Tier 3 (Correlational):** Correlational studies can be useful to look for "no harm" and for suggestive evidence of benefit. For example, educational leaders can use correlational evidence to examine whether the intensive use of a new AI-based product is associated with negative or positive outcomes, both for students overall and within specific student groups. Likewise, correlational studies can reveal whether using a product for a recommended or longer period of time is more strongly associated with benefits than using the product for a minimal amount of time. Positive correlation between the recommended usage and desired outcomes suggests benefit.
- **Tiers 1 and 2 (Quasi-Experimental and Experimental):** Studies at these levels compare a new approach (e.g., using AI) to an existing approach, either by randomizing students to the two approaches (Tier 1) or comparing well-matched students who use the alternative approaches (Tier 2). Although these studies have a reputation for being costly and difficult to conduct, faster and less-expensive options are increasingly common. For example, a district may have adopted a literacy solution in its non-AI form, and the literacy product may have a new version with an AI-based chatbot. It may be feasible to conduct an A/B comparison with and without the chatbot at low cost and modest complexity. Educational leaders may prefer to wait for rigorous Tiers 1 or 2 evidence to become available over time before adopting AI-enabled products.

Discussion Points and Questions

These discussion points and questions are intended to catalyze conversations among your learning community regarding Evidence.

Discussion Points:

- Discuss why using evidence is important and how your educational organization presently uses evidence in decision making.
- Share ways in which your leadership team can learn more about how to use evidence to support edtech adoption.
- Share evidence for AI-enabled products that are being considered for adoption or purchase and evaluate the quality of this evidence, identifying what information is missing/desirable for making a sound decision.
- Consider how you might collaborate with researchers (e.g., from a local university or nonprofit) to evaluate evidence for AI-enabled products.

Discussion Questions:

- What are your educational leadership team's requirements for evidence for different ways of using AI (and all edtech) in your setting?
- What low-cost and timely evidence gathering activities can be integrated into existing procurement processes to support evidence-based edtech procurement decisions?
- What steps can be taken to move toward building a stronger evidence base for use of AI in schools, including not only efficacy on average but also avoiding potential harm to any student?

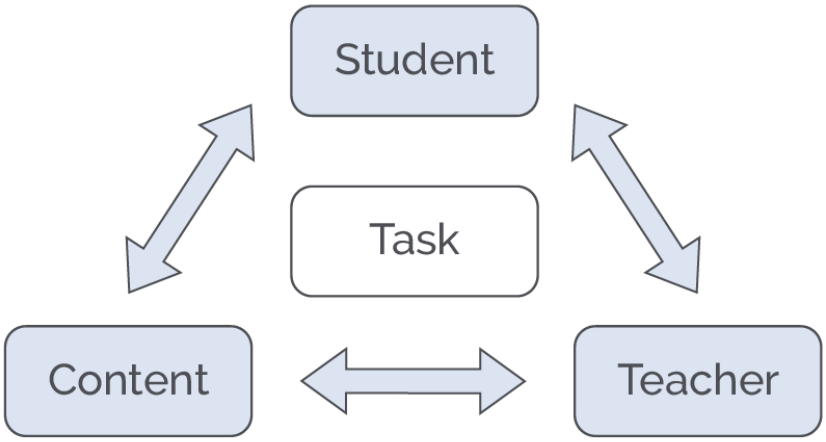
Module 5: Considering the Instructional Core

Defining the Instructional Core

Richard Elmore (2008) defines the instructional core as the connection between students, teachers, and content, grounded in high-quality instructional tasks. To improve student learning at scale, it is essential to address the instructional core by focusing on three key principles:

1. raising the level of content taught to **students**
2. enhancing **teachers**' skills and knowledge related to that content
3. increasing students' active engagement with the **material**

Together, these efforts form a strong triangle that can drive meaningful improvements in education outcomes.



The concept of the "instructional core" pinpoints the difference between innovations that deeply advance student learning and those that merely scratch the surface of educational improvement. In another foundational article, Cohen and Ball (1999) write: "Since World War II, efforts to improve schools have numbered in the thousands…. Unfortunately, three decades of research has found that only a few interventions have had detectable effects on instruction and that, when such effects are detected, they rarely are sustained over time" (p.1). They propose

that meaningful improvements require reconfiguring instruction, which they define as the interactions among teachers, students and curricular resources. For students, this means they are central to the learning process, actively engaging with challenging content and meaningfully engaging with skilled, knowledgeable educators. For teachers, this means enhancing their expertise, pedagogical strategies, and instructional practices in order to effectively teach and engage students and help them learn and master rigorous content. For educational leaders, this means providing professional learning opportunities, tools, and targeted mentoring that specifically advance teachers' instructional skill and capacity and impact the level of student work in classrooms.

Multiple states and school districts have centered the instructional core in their work. For example, Nebraska's approach observes as a first principle: "Increases in student learning occur only as a consequence of improvements in the level of content, teachers' knowledge and skill, and student engagement." And Kentucky provides Characteristics of Highly Effective Teaching and Learning, a shared framework for discussing best practices for creating ideal learning environments.

The Instructional Core and AI

The Department sees the potential for effective use of edtech to enhance and strengthen the instructional core. There are, of course, multiple examples of the use of AI-enhanced tools or applications that fall short of meeting this potential. For example, some applications of AI generate relevant or attractive content for lesson planning but might not be accurate or improve the quality of the content. Some applications of AI can save teachers' time but fall short of providing them new knowledge, skills, or strategies that will enhance their interactions with students in scaffolding robust content. And as every teacher knows, technology can sometimes be used in a manner that increases students' engagement superficially without promoting the skills and growth mindset that students need to exercise critical thinking and reasoning and meaningfully exert effort and persistence around an instructional task.

To apply AI with purpose, educational leaders must be attentive to avoid superficial or inconsistent applications of AI and instead maintain a focus on how AI can be utilized as a tool to improve the quality of teaching and learning, student-teacher interactions, and instructional tasks at the heart of the instructional core. For example, in teaching reading, the National Center on Improving Literacy (2022) argues that by integrating comprehension with decoding while **building on students' prior knowledge**, educators can accelerate reading growth. An effective use of GenAI might **revitalize the content students are asked to read** while also enhancing **how students and teachers interact** while they read. In mathematics, students need practice using concepts to explain mathematical ideas, but too often many mathematics assignments remain focused on practicing low-level procedures (TNTP, n.d.). An effective use of GenAI might **provide students more opportunities to practice explaining** core mathematical concepts, integrating both student feedback and supplemental resources for teachers to help them develop expertise in connecting student reasoning with mathematical concepts. Next Generation Science Standards (NGSS Lead States, 2013) set expectations that classrooms will

increase attention to "cross-cutting concepts," which connect individual units of study in the core subject of science. An effective use of GenAI might connect the units of study to broader scientific concepts and other academic subjects, like social studies or math.

Centering the instructional core requires interconnected changes, some of which reach beyond the classroom triangle in which teachers, students, and content interact. Assessment is a huge factor in what changes or stays the same related to instruction. The effective use of AI can transform how teachers and students make use of assessment results for instructional planning, decision-making, and the implementation of effective instructional interventions. In our earlier AI Report, the Department observed the opportunities for AI to enable measuring what matters while also observing that the assessment community has strengths in addressing fairness and mitigating bias that will need to be brought to bear on AI-enabled assessment. Educational leaders play an essential role in supporting classroom **educators to make the right connections** between assessment systems and improvements to the instructional core.

The Department's AI Report notes that many educator preparation programs attend to technology in superficial or inconsistent ways. Of course, this also pertains to in-service teacher professional development. For educators to integrate AI into the instructional core in their classrooms, they will need more time and higher-quality professional learning designed to develop and enhance their expertise with using edtech. Educational leaders can influence change by asking hard questions about whether and how AI literacy for teachers can help advance the instructional core.

Finally, the AI report emphasizes the potential of AI to reduce the administrative burden on teachers, giving them more space to focus on their students and on instruction. The potential to make teachers' lives better can be an initial motivator to explore the use of AI; however, teachers' aspirations will quickly seek to expand the potential for AI to support effective instruction, beyond enhancing time management and productivity. Educational leaders have a role in keeping a focus on advancing the instructional core as additional time or capacity becomes available.

Discussion Points and Questions

These discussion points and questions are intended to catalyze conversations that connect the opportunities surrounding AI to local efforts to improve the instructional core.

Discussion Points:

- Discuss how AI-enhanced products complement the instructional core as a catalyst.
- What are "higher level" or better-quality tasks that students, with the use of AI-enabled tools, can successfully do?
- If AI supports students in some interactions as they work on these tasks, what are the most valuable ways teachers should interact with students?
- How can teachers learn new knowledge, skills, and instructional strategies using AI?
- Share ways in which your leadership team can connect current or potential uses of emerging technologies with its priority on the instructional core.

- Create guidance for selecting AI-enhanced products that are not pure supplements to instruction, but rather support desired instructional changes among teachers, students, and resources.
- Explore the additional factors in your setting – beyond students, teachers, and resources (e.g., data from edtech tools and digital platforms and the interpretation and analysis of those data to answer administrative questions) – that most strongly influence the instructional core and how AI might enhance these factors or pose new risks.
- How might you increase reflection and action to create the conditions necessary to harness the potential of AI?

Discussion Questions:

- Imagine an observation in a classroom that is using AI. What does success look like?
- How are students engaged in tasks, what tasks are they doing, what are they doing and saying?
- How are students supporting each other?
- What are teachers doing and how are the knowledge and the level of skills they are bringing to the instructional process improving?
- What are their families saying and noticing?
- What guidance and examples (Module 9) are your educational leadership team providing to ensure coherence among the different ways of using AI (and all edtech) as it relates to the instructional core?
- How can staff development and collaborative spaces prioritize interconnectedness between instructional core and AI, rather than seeing these as separate issues?
- When one element of the instructional core is changed, how are you balancing the others to ensure consistency?
- What cost-effective and timely professional learning experiences can support your team as they prepare to leverage AI-enhanced tools in their practice?
- How are educational leaders examining the comprehensiveness of teachers' unique roles and opportunities to learn?
- How are supervising teachers receiving additional support to coach and assist student teachers on their use of AI in relationship to the instructional core?
- How can teams in your educational setting use research to understand the differences between promising uses of AI-enhanced tools and those which are likely to fail to meaningfully change the instructional core? Further, how can research and staff expertise enable consideration of equity divides (e.g., in the NETP) throughout this process?

Module 6: Planning Your AI Strategy

Inform and Listen

The Department, through listening sessions, learned that educational leaders are experiencing wide variation in their communities' knowledge about AI, their excitement and concerns, and their views on important next steps.

As a first step, educational leaders can host a listening session in their own communities to engage a variety of voices and involve invested community members in developing guidance and recommendations. As discussed in the Evidence module (Module 4), the experiences of an educational community are also evidence that should be weighed while developing a strategy for responsible use of AI. Module 5 provides insights into the process of hosting a **virtual** listening session. A virtual listening session commonly lasts 60-90 minutes, and it is advisable to hold multiple sessions for various audiences within your educational community.

Preparing for the Listening Session

The Department anticipates that individual educational leaders have their own best practices for planning meetings, including, for example, what format to use (virtual, physical, or hybrid), who to invite and how to invite them, what accommodations to provide for those with special needs, and how to capture notes. As many educational leaders already know, an event agenda can be a useful organizing document for everyone who will participate in conducting the event.

Some specific considerations for a listening session about AI may include:

- **Local experts:** Consider featuring leaders, educators, students, or community members that represent the various demographics of the school ecosystem who have knowledge, expertise, and the ability to talk about AI in terms that are likely to be clear and accessible to those who will be participating.
- **Concrete examples:** Consider inviting educators or students who can provide examples of how they have used AI in responsible ways, or conversely, establishing clear limits on when AI should not be used.
- **Balanced panelists:** Consider whether a panel might be able to set a responsible tone for discussing the range of opportunities and risks with AI, and thus model productive discourse about AI. This panel should include a range of people that best represent the school ecosystem. This might include, for example, those who can share a family perspective, those who identify as having a disability, and out-of-school-time programs who work closely with the district.

In addition, educational leaders may wish to consider:

- Who is the **audience** for this listening session?
- Should the listening session be **invitation only or held as a town hall**?
- What **information** about AI, edtech, and education will the session leader share and receive during this session?
- Who can serve as **trustworthy** and **knowledgeable** facilitators, panelists, or speakers to engage this audience? What is their connection to the audience?
- What **questions or concerns** should the organizers collect before the listening session so that leadership can be better prepared?
- How can the organizers facilitate **interaction** while managing challenges that could arise during a session? What are the various modes of engagement that the organizers should consider?

During the Listening Session

The listening session should prioritize hearing from district and community members. Educational leaders can set expectations for the participatory aspects of a listening session via an opening presentation. In planning the opening presentation, consider how the presentation addresses the following considerations:

- **Purpose:** State why you are holding the listening session; one example might be to engage the community in building a strategy to guide both current and future uses of AI in schools. Describe what the district has already identified as key principles, values, guidelines, or commitments. Explain how insights from the listening session will be used.
- **What is AI?** Establish some common ground in the audience's experiences of AI (for example, automated maps or phone-based voice assistants), accessible definitions (for example, "automating based on associations" from the AI Report), and clear examples. Share what kinds of uses of AI are already occurring in the school community.
- **Where is the district in the development of a strategic process?** For example, what policies, guidelines, and guardrails already exist (for AI, edtech, and non-technology issues)? On which principles or values is there agreement? What are some of the new challenges?

Educational leaders should keep the following in mind when hearing from the audience:

- **Responses:** How are community members responding to the integration of AI? It can be helpful, for example, to acknowledge the spectrum of concern-to-enthusiasm that participants may bring to the listening session.
- **Moderated interaction:** Consider whether accepting questions and comments via notecards, an electronic poll, or other means might be preferable to allowing anyone to speak on a live microphone.
- **Call to action:** Help participants understand how they can further contribute to solutions. Explain how their feedback will be applied and how they can continue to provide guidance.

Here is an example of a planning document that you can use and share with all speakers and members of the planning team.

Outline materials in an event agenda and share the document with all speakers for visibility.

After the Listening Session

Audience members should not leave a listening session without follow-up, especially after sharing their comments and experiences. Educational leaders might consider the following to acknowledge that the audience's time and input is valued and will be used for future development:

- **Send a thank you message** to attendees.
- **Summarize** the session, potentially using an AI tool to help identify themes that are then reviewed by humans.
- **Share** a brief recap of key insights and lessons learned from the listening session with attendees. Include opportunities for continued engagement and other next steps.

Pace is a community agreement often dictated by consensus and governed by the needs of a team's most vulnerable members. Both in its listening sessions and at public conferences, the Department learned of educational leaders' concerns with the pace of AI advances. Furthermore, educational leaders, parents, educators, and students all expressed reservations about how past technology breakthroughs have impacted their communities. In light of past experiences with previous edtech integration, educational leaders in listening sessions expressed desires to set priorities and to regulate the pace of adoption of newly emerging technologies. Rather than "letting a thousand flowers bloom" by allowing students and educators to introduce whatever applications seems attractive, leaders seek a more focused and strategic approach to integrating AI into their settings.

The Department encourages educational leaders to intentionally set priorities according to a shared vision for education (as discussed in the second recommendation in the AI Report). Further, the Department encourages educational leaders to establish a pace and support their community in saying "no" or "go slow" as conditions warrant.

Planning an Approach to Prioritization

GenAI is being used in every type of edtech and for wide-ranging educational applications (e.g., teaching and learning, assessment, counseling, school safety, logistics, operations, and more), and the market is moving at a blistering pace. Because of this rapid movement, priorities and pace should not come from the technology itself.

In its AI Report, the Department recommended starting from a shared educational vision, which is likely stated in an educational institution's mission, strategy, portrait of a graduate, and other documents. In addition, many states, including Washington and Delaware, are now producing guidance specific to AI. The nonprofit sector is providing additional input. Reviewing the three introductory modules is also a good starting point to ground planning in a strong foundation that mitigates the risks of AI in education described in Section 1 of this toolkit.

Each educational institution will likely need to develop its own approach to prioritization based on how it expresses its mission, vision, strategy, and other key input.

Discussion Questions

Educational leaders can ask:

- What are our most important unmet educational goals that AI can help resolve?
- Where are we starting in terms of responsible use of edtech?

- What has worked previously in terms of a pace for responsible adoption and effective technology implementation?
- How do we ensure attention to important problems that do not have immediate technological solutions (e.g., truancy) with attention to problems that are a fit for AI and other emerging technologies?
- What is our human resource capacity and readiness to implement AI in the best possible ways?

Specifying Factors

The Department recommends that educational leaders list key topics on which their teams will gather information before making decisions about priority and pace. A partial list of considerations could include:

- **Alignment to Educational Strategy:** How tightly aligned is this opportunity to use AI to the educational institution's vision, mission, strategy, and goals?
- **Risk Analysis:** Considering the areas of concern described in Modules 1, 2 and 4, what are the risks with this use of AI in our educational setting, and to what extent are those risks managed?
- **Content Moderation:** To what extent will the AI tool impact student learning due to the data used to train the underlying model?
- **Evidence:** What evidence (see discussion in Module 4) supports prioritizing this opportunity to use AI?
- **Capacity and Readiness:** To what extent can we engage the necessary leadership, resources, and talents in our educational community to support success?

Educational leaders may wish to consider four or more levels of adoption such as:	
Encouraged Use	Some proposed uses of AI emerge as high priority given the factors that were determined to be important in the educational setting, and may be encouraged.
Allowed Use	Some proposed uses of AI may be permitted, but not especially e encouraged, and leaders may wish to monitor how much use occurs and what the impacts are.
Limited Use	Some proposed uses of AI may be worth exploring in a bounded and closely monitored way.
No Use	Some proposed uses of AI will be unrelated to important educational goals, too risky, lacking evidence, or too hard to implement well.

Educational leaders should set expectations on pacing for how soon and how much use of AI is expected, for example, compared to existing resources and approaches. **Likewise, allowing enough time to evaluate, reflect, and adjust plans for AI use is important.** Pace may also be reflected in how many AI opportunities are allowed or encouraged.

Module 7: Guide and Support

In implementing a strategy to integrate AI in your school, there will inevitably be situations where the challenges faced exceed present capabilities. In such cases, seeking external assistance may become necessary. This paradigm shift into the GenAI era is a similar venture that may require assistance to get started, to maintain, and to complete successfully. As is widely reported in educational news media, educators, students, parents and caregivers, community members, and others in learning communities are asking for guidance about responsible AI use. States, districts, nonprofits, and other organizations are responding. Building on their recommendations, the Department urges educational leaders to establish a task force in their institution (or another similarly chartered council, committee, or team) to respond on an ongoing basis. A task force can provide more specific guidance and support to educators and others about responsible uses of AI that are well aligned to the institution's priorities.

Planning a Task Force

The Department expects the needs for support and guidance to vary across settings, and thus there is no single charter for a task force that will work in all settings. Questions to ask include:

- **Who will be included in the task force?** The process of choosing task force members will depend on the resources available in individual districts. No matter how they are chosen, it is important to include diverse experiences and voices. This can either be accomplished through an open call for volunteers, applications from interested community members, or strategic selections to ensure diversity. If the latter, the selection process must be transparent, consistent, and fair. Members of the task force must have an understanding of the particular risks that might arise in the use of AI, as identified in previous modules (see the Privacy and Civil Rights modules).
- **Can the task force be integrated into existing district structures?** Some districts may not have the resources for a separate task force; also, interested members in a task force may not have the time and bandwidth to join an additional group. Districts could explore existing spaces where the task force can be added, such as district academic support teams.
- **How will task force members be compensated?** It is not easy for educators, students, and community members to add more time to busy schedules to attend task force meetings and accomplish task force duties. If districts consider compensation in some form, it could, for example, be monetary, include extra credit for students, or release educators from other responsibilities.
- **Who will the task force support and guide?** A task force might be chartered mainly to support educators, or to help building-level decision makers, or to work with students and families, or some combination of these groups.
- **What will be the task force's charge?** A task force might help educators implement an acceptable use policy, given emerging applications of AI, or to address equity divides (see Module 3), or to vet possible applications of AI for broader adoption and implementation.

- **What knowledge will the task force use?** A task force will need input on the educational institution's desired pace and priorities, as well as information garnered from listening sessions, and guidance from leadership.
- **What resources will be available to the task force?** How much time will they have to meet and work? Will they be able to conduct surveys? Will they be able to go to conferences or attend professional development trainings to increase their own knowledge about AI? Who will help the task force gather evidence from research?
- **How will the task force provide support and guidance?** Will members of the task force each work with members of the educational community in their specific building or team? Will the task force provide professional learning to educators? Will they conduct awareness or other campaigns? Or will they provide recommendations to others who have responsibility for support and guidance?

Ethics as a Cornerstone

To build public trust and ensure that AI tools are integrated in ways that respect individual rights, promote fairness, and prevent potential harm, an AI task force needs to intentionally center ethics. The Department suggests educational leaders provide ethics training to a task force, framing the ethical adoption, application, and use of AI as a cornerstone of any school's or district's AI strategy. This includes familiarizing task force members with broadly applicable ethics concepts and frameworks as well as those specific to AI and education. Then, task force members should make explicit the ethical considerations they want to hold at the center of their process. A shared ethical purpose is critical for the complex work of adopting AI in learning communities.

In February 2024, researchers at NIST suggested building on long-standing concepts in the 1979 Belmont Report, which provides and develops guidelines on the basic ethical principles that should underlie the conduct of human research. These long-standing concepts include beneficence, respect for persons, and justice, which can be used to organize an approach to ethics in the age of AI. Researchers and educators have specifically been working together to develop ethical guidelines, and the Department expects this work to continue.

Additionally, the Department's review of the variety of ethics frameworks for implementing AI solutions, including the 2024 NIST AI Risk Management Framework, found that general ethics and AI ethics concepts apply equally to education: equity, sustainability, transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom, and autonomy.

Value-Centered Design Ethics in Education	
General Ethics Themes	**Education Ethics Themes**
• Transparency • Justice and fairness • Non-maleficence • Responsibility • Privacy • Beneficence • Freedom and autonomy	• Pedagogical appropriateness • Children's rights • AI literacy • Teacher well-being • Equitable instruction & access

Beyond adapting generally accepted ethics themes to more closely fit education settings, the Department's review of the 2024 NIST AI Risk Management Framework identified four additional principles specific to education: pedagogical appropriateness, children's rights, AI literacy, and educator well-being. Student well-being was not specifically listed in the review; nonetheless, the Department finds student well-being to be of paramount importance. These ethical principles are core components of designing AI systems that interact with children.

Creating a Charter

The Department found that freely available GenAI tools can create an initial outline for a task force charter. Below is a sample charter that was first drafted using Google's Gemini, and then substantially modified by the authoring team. The Department chose to include this example to illustrate the steps that educators may take when using AI to help with their work. The detailed content within the example is not directly from GenAI but reflects additional edited content and expertise from the authors.

Box: Example of an Artificial Intelligence (AI) in Education Task Force Charter

I. Introduction

The [School District Name] School Board recognizes the growing potential of Artificial Intelligence (AI) to create opportunities to improve education but also to introduce both anticipated and unanticipated risks. To ensure responsible and effective integration of AI tools, the School Board hereby establishes the AI in Education Task Force.

II. Mission

The mission of the AI in Education Task Force is to document the opportunities and risks of AI in our district. The Task Force will develop a plan for the ethical, equitable, and sustainable use of AI to enhance student learning (and may later address support for educators and contributions to district operations). The Task Force will respond to issues that arise as students and educators use AI for learning.

III. Goals

- Identify current and potential uses of AI for student learning in the district.
- Gather evidence on the uses both from published research and from experience in the district.
- Collect information on topics related to the risks of these uses of AI including data privacy, algorithmic bias, content moderation that impacts learning, and equity risks.
- Create an initial risk management plan based on the information collected in the previous bullet point.
- Recommend professional development opportunities for educators and staff regarding using AI to support student learning.
- Create a communication plan to keep our community informed about the Task Force's work.

IV. Membership

The Task Force will be composed of a diverse group from our community, including:

- District Administrators
- Curriculum Specialists
- Technology Experts
- Educators (representing various grade levels and subjects)
- Students
- Families, Caregivers, and Community Representatives

The Task Force may invite additional members with specific expertise as needed.

V. Meetings and Deliverables

- The Task Force will meet [frequency] for a period of [duration].
- Meetings will be open to the public (optional, depending on meeting format).
- Minutes will be recorded and posted publicly (with sensitive information redacted if necessary).
- The Task Force will produce these deliverables:
  - An inventory or classification of the uses of GenAI for student learning in our district
  - An analysis of what quality of evidence for these uses exists and what kinds of evidence is most strongly needed

- A discussion of risks and an initial risk management plan
- Recommendations for professional development and other next steps

VI. Support and Resources

The School Board will provide the Task Force with the necessary resources to fulfill its mission, including:

- Staff support
- Meeting space
- Compensation to members in the form of [...]
- Access to relevant data and research
- Budget for potential expert consultation

VII. Review and Revision

This charter may be reviewed and revised at any time by the Task Force and with the approval of the School Board.

VIII. Conclusion

By harnessing the power of AI responsibly, we aim to create a more engaging and effective learning environment for all students.

Maximizing Opportunity: Guiding the Effective Use and Evaluation of AI

This section addresses the implementation and use of AI in schools. Following the development of a strategy and plan for pacing and introducing AI to a school community, these three modules provide guidance on key actions that educational leaders can take when using, managing, evaluating, and scaling AI in their districts and schools.

- Module 8 supports educational leaders to **build AI literacy for educators.**
- Module 9 discusses top level considerations for educational leaders **updating their policies for acceptable use of technology.**
- Module 10 suggests how educational leaders can create an **organization-wide action plan** for implementing AI in their educational communities.

Each module includes corresponding scenarios in Appendix 3 for educational leaders to practice applying what they have learned in real-world contexts. An educational leader who is reading this section can ask: What key ideas should be included in AI literacy, acceptable use, and risk management policies?

Module 8: Building AI Literacy for Educators

The responsible use of AI requires building a strong foundation and knowledge of AI literacy. This includes the knowledge, skills, and attitudes needed to engage with AI safely and effectively.

AI literacy for educators was a top recommendation from educators in the Department's listening sessions and is also a priority for leaders across Federal, state, and local levels.

Developing AI literacy is essential both for students and for educators. This toolkit recommends prioritizing AI literacy and capacity first for educators, then empowering them to do the same

for students and others in their school communities. Developing educator AI literacy is important because:

- With greater knowledge about what AI is, how it works, and which uses of AI are recommended in classrooms, educators can make **informed decisions** about how to integrate AI in their classrooms.
- With greater awareness of how issues of ethics and risks arise when using AI, educators can **advocate for safety** among their students and broader communities and undertake their role as humans-in-the-loop as AI becomes more prevalent in their students' experiences.
- With a forum to develop understanding of how AI may impact the workforce, and the world students will enter, educators can better **guide students toward successful futures in a world where change is occurring at a fast pace**.

Defining AI Literacy

AI literacy includes the knowledge, skills, and attitudes needed to engage with AI safely. (Mills et al., 2024). This skillset is about understanding AI's strengths, limits, and impacts to make informed decisions about its integration and to prepare learners for the AI-driven future (Kulesa et al., 2024).

As an AI literacy initiative progresses, all educators should be able to **actively recognize** AI in multiple forms, make sense of their own **interactions with AI**, make informed decisions based on an **understanding of how AI is built** and its limitations, and view AI through a lens that includes **ethics and social impact**.

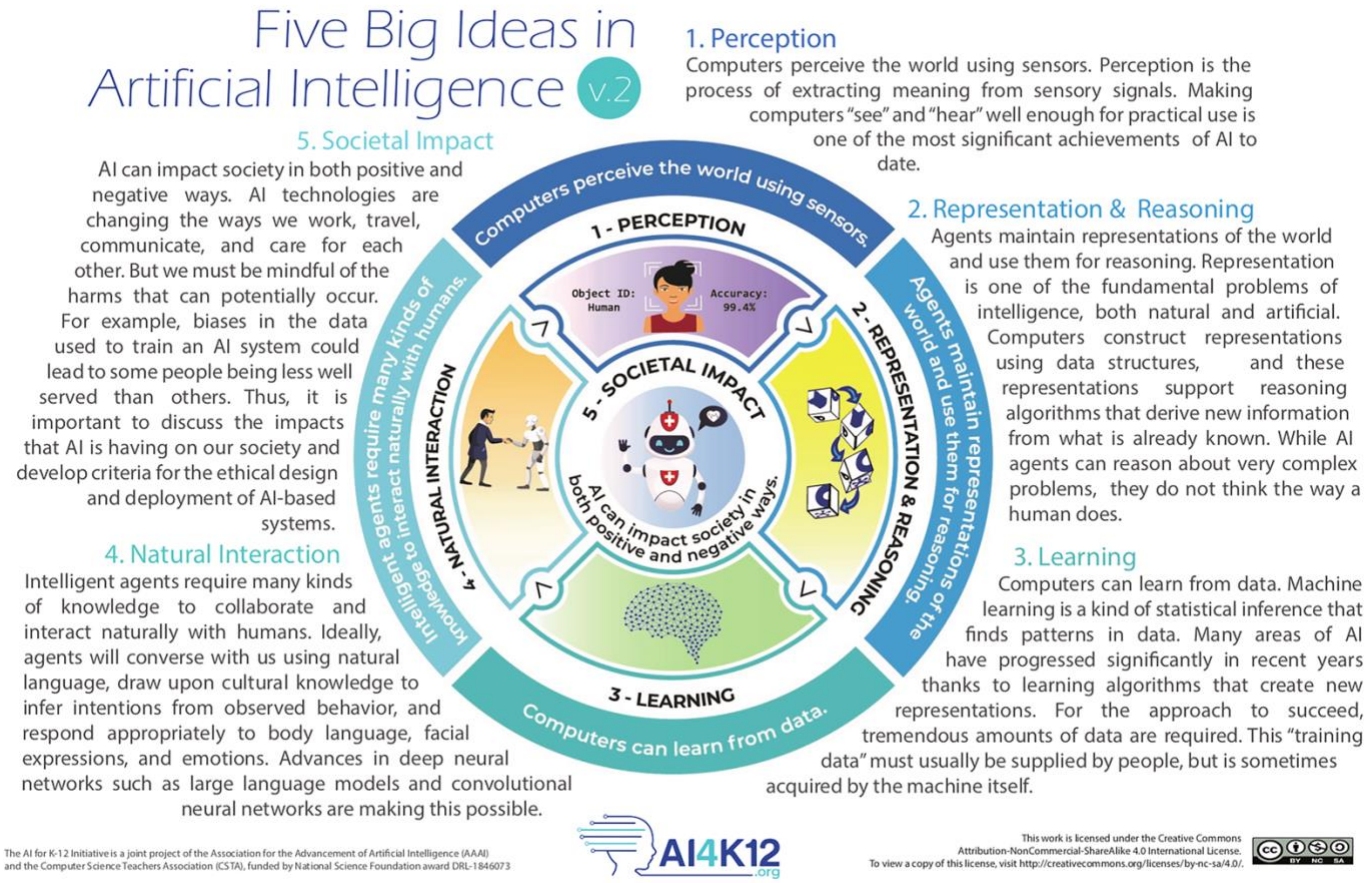
Key activities to address in AI literacy programs include **Understanding**, **Evaluating**, and **Integrating into the Classroom.** Each activity is discussed in a section below. Educational leaders are expected to make their own decisions, based on their specific context, about how to best cover the breadth of content over time. Activity 1: Understanding AI

The above definition of AI literacy calls for comprehending a wide variety of topics. Based on our review of recommendations in the field (Druga et al., 2021; MIT STEP Lab, 2024; World Economic Forum, 2024) as well as the content of the National AI Literacy Day, the Department recommends that AI literacy initiatives cover the following topics. These topics are best delivered over time to give educators time to develop understanding and may be delivered in any order.

- **Defining AI.** In its AI Report, the Department suggested conceptualizing modern AI as "automations based on associations" and critically discussed other definitions that rely on analogies to how people think—because machines generate outputs in different ways than people do. Many definitions of AI are now available, including the Executive Order on AI, which defines AI as "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments." The Department recommends looking for AI literacy offerings

and facilitators who can work with educators to critically examine a variety of definitions, and who can work with educators to demystify the fundamentals of AI.

- **AI's History and Origins.** Educators should be aware that AI's history extends back to the earliest days of computers in the 1940s (e.g., the work of Alan Turing), and that humanistic and ethical issues have accompanied progress in the field throughout. AI arose from mathematics and computer science and has moved through a series of approaches that precede today's GenAI approaches. The educational use of AI dates back to 1970, and many prior forms of AI have resulted in useful educational applications (Wooldridge, 2021). The Department recommends looking for AI literacy offerings that provide a sense of history, both for AI and for AI's prior uses in education.
- **Data and Machine Learning.** Modern AI is engineered by identifying patterns in large volumes of data, by a set of techniques broadly called "machine learning." Although educators may not need to understand the differences among machine learning algorithms, they should understand the associative, statistical, and probabilistic nature of machine learning. Doing so will help educators understand why AI sometimes produces erroneous outputs. Without understanding the likelihood of generating erroneous outputs, users may utilize AI-generated information with no review, which can lead to incomplete understanding or false information being learned. Likewise, understanding the process of using data to build AI models can inform educators about the human role in decision making at each stage. Steps such as collecting, cleaning, curating, labeling, analyzing, and evaluating can be presented in ways that build on educators' existing knowledge. Such transparency can support educators to understand how and where biases and inaccuracy can arise.
- **Breadth of AI Capabilities.** Educators should be aware AI is not just chatbots, nor is it confined to textual input and output. Indeed, many experts see "AI" as an umbrella term that encompasses different subfields with distinctive approaches. One way to describe the different aspects of AI is the Five Big Ideas image.



These and other aspects of AI are rapidly being incorporated in a wide range of educational products and services. Educators are likely already familiar with a range of AI applications, such as mapping, robotics, voice assistants, self-driving cars, recommendation systems and more. The Department recommends increasing educator awareness of the range of AI applications and their current use in real life, so they are prepared for the future.

- **Existing and Emerging Uses of AI in Education**. AI literacy should build educators' knowledge of types of algorithms, automations and AI techniques that have been used in education or may be used in the future. One decades-old example is the use of intelligent tutoring systems to support students as they practice mathematics. Algorithms also have been used to identify students who are "at risk" and to score essays for decades. In adaptive learning systems, algorithms determine what content is assigned to students and what feedback they receive. Educators will benefit from coverage of examples with longer histories (like these), examples that are appearing now (such as uses of Chatbots), and future-oriented examples (e.g., where educators or students may use AI to create movies or interactive game-like experiences to pursue learning objectives). It would be beneficial to accompany examples like these with discussion of the types of machine learning or AI used in the application, with some attention both to the capabilities of the technology and its limits or risks.

- **Ethics and Social Impact.** Building on such examples, educators should consider how ethical concepts such as fairness, accountability, and transparency apply to the ways that educators and students select and use AI tools. It is important to acknowledge the potential risks of bias and discrimination that AI technologies may carry, particularly against historically marginalized groups. Understanding these risks and their implications for civil rights is critical for designing equitable learning environments. Students and communities often bring up concerns with the social impact of new technologies, including issues related to bias, fairness, and discrimination, and AI is expected to have broad impacts for the workforce, digital citizenship, and in everyday life. The Department recommends that AI literacy offerings provide ample time to discuss ethics and social impacts, both as they impact educators personally and as they shape the world their students currently live in and will enter in the future.

As an AI literacy initiative progresses, all educators should be able to actively recognize AI in multiple forms, make sense of their own interactions with AI, make informed decisions based on an understanding of how AI is built and its limitations, and view AI through a lens that includes ethics and social impact.

Activity 2: Evaluating AI

Educators will often need to evaluate current or proposed uses of AI, which includes both formal school adoption and informal individual choices by educators and students to use what is available to them for educational purposes. Building on the understanding developed in the prior section, AI literacy should prepare educators to evaluate AI uses in educational settings at multiple levels.

The uses of AI in a district setting are likely to be broad. For example, the Department's AI Report contains chapters on adaptive learning, supporting educators, and improving assessments. AI may also be used in logistics and operations such as lesson planning. An AI literacy curriculum should prepare educators for the potential breadth of use, and not be limited to narrowly evaluating chatbots or any other specific form of AI.

Earlier modules in this toolkit also support educators in evaluating AI. The Department recommends outlining how educators should evaluate AI by considering the following factors previously addressed:

- **Privacy and Data Security**. AI literacy can build on what educators already know about privacy and data security, but also should include attention to new or intensified challenges such as issues related to collecting and using biometric data.
- **Civil Rights and Digital Equity.** This topic encompasses protections from bias and algorithmic discrimination, as well as ways in which AI may better support access and learning for all students. Evaluating AI's capabilities to support diverse learners should also consider issues in the National Ed Tech Plan corresponding to the access, use and design divides.

- **Evidence.** AI literacy should reinforce the importance of looking for and considering evidence as technology is evaluated for educational use. Types of evidence relating to rationale, efficacy and safety are discussed in the earlier module.
- **Alignment to Opportunities and Mitigation of Risks.** AI literacy should emphasize choosing uses of AI that align to an overall educational vision and strategy and avoiding being distracted by misaligned alternatives. Ethics should also be considered during evaluation. Module 4 lists additional risks.

The table below provides some guidance for evaluating edtech. As an exercise, you can also read the scenarios and consider how you would react in these situations.

LEARNER-CENTERED EMERGING EDTECH EVALUATION RESOURCES	
Is there reliable evidence that the AI-based product or service works?	• What Works Clearinghouse • Kentucky Department of Education Evidence-based Practices • The EdTech Evidence Exchange • Ed Tech Adoption
How does the AI-based product **promote equity and protect students' civil rights** in its design and on-going support?	• OCR Reading Room • The CRDC's Data on Equal Access to Education • CASEL Equity Resources • Inclusive Intelligence: Impact of AI • Infrastructure in Inclusive Technology Systems
What **safety measures** are incorporated into the AI-based product and how do those align with the school's policies and requirements?	• U.S. Department of Education's Student Privacy Policy Office • NASSP's Student Data Privacy • The Student Data Privacy Consortium
Do reviews and oversight from trusted parties support the AI-based product's value and trustworthiness?	• Common Sense Education • ISTE/EdSurge Product Index • Gartner Education • G2 Education • Capterra Education

Activity 3: Integrating AI into Classrooms

Local educators and educational institutions make decisions about which applications of AI to integrate into classrooms and will design professional development experiences accordingly. This section focuses on five broader considerations in planning how an AI literacy initiative addresses classroom integration.

1. **Balancing immediate implementations vs. time for collaborative planning and support.** The Department's AI Report observed a tension between racing to implement AI in classrooms and taking the time to understand context and create the right conditions for implementation. With regard to getting the context right, there are two issues: (a) uncovering specific factors in the classroom that may contribute to success or failure and (b) creating the right surrounding conditions for success. Likewise, Module 6 strongly encourages educational leaders to establish a sense of pace and priority, so that AI implementation is not rushed. AI literacy should give educators hands-on opportunities to explore AI applications, while also providing time and space for discussions with fellow educators, coaches, and support staff about how to best integrate AI into their specific settings. Educators themselves are in the best position to appreciate how to connect the specifics of their classroom context to new technological approaches but need time to work with each other without the expectation that they will immediately translate an offering into classroom use.
2. **Balancing productivity uses and transformative uses.** Supporting classroom teachers through effective edtech should strike a balance between "doing things better" (productivity) and "doing better things" (transformation). AI could increase productivity on routine tasks (such as writing a first draft of an email to a parent) so that educators can spend more time with students. AI also might more deeply transform educational practice, for example, by revitalizing the process of formative assessment or by generating lesson plans and unique content on demand that more closely reflect their students' backgrounds, yielding increased engagement. Using AI in the most beneficial ways may require pedagogical change, not just pedagogical efficiency.
3. **Establishing student and educator agency.** Students and educators need guidance and access to learning experiences that help them develop and practice AI literacy skills. This practice will help them improve their ability to understand when and how to implement AI tools for school-related activities, including tool selection and evaluation. Without a deep understanding of the three modes of engagement (understand, use, and evaluate), students may rely on AI tools and trust their outputs without critical examination, which could impact their access to information and learning experiences. Additionally, if educators use AI tools to monitor student use of AI tools, students may interpret this as mistrust, which may reduce their willingness to engage in classroom activities.
4. **Developing an understanding of human bias and the strong potential for AI bias.** Educational leaders should provide guidance to educators to help them recognize and acknowledge their existing cultural and societal biases, especially regarding how students from historically marginalized backgrounds are disproportionately disciplined around

biased views about academic integrity, and how using AI tools to detect disingenuous student use can perpetuate and worsen said disciplinary inequities. These supports can come in the form of guidance in responsible use policies, professional learning tools, and enhancing existing equity trainings.

5. **Establishing feedback loops that lead to improved use of AI over time.** In a simple feedback loop, AI literacy should help educators recognize any shortcomings of the initial output they receive (such as a lesson plan) and improve the output before using it. A more systematic feedback loop would be a "plan-do-reflect-improve" cycle (or a similar variant). Educators often meet with each other repeatedly to figure out what's working and how to make improvements. Educators are also likely to be in feedback loops with educational leadership to discuss how to align opportunities and mitigate risks.

Educational leaders must address current educator needs and plan for successful long-term implementation of AI in classrooms. Thus, educational leaders may ask: **How can I help educators develop the necessary skills and practices to improve the integration of AI into classrooms in ways that meet the specific goals and needs of this learning community?**

Put it into practice! How will your district react in situations that require applying AI literacy skills?

See Appendix 3 for scenarios to review and share with your team.

Additional Considerations for AI Literacy

The three activities discussed above can organize the content of an educational leader's AI literacy initiative. In the course of reviewing input from the Department's listening sessions and publicly available documents about AI literacy, the Department also recognizes cross-cutting considerations, which are shared below:

Knowledge Considerations

Demystify AI	Explain how AI works and why it sometimes produces biased, wrong, or risky outputs. Use minimal jargon and find relatable insights to share.

Make Connections	Make connections and build on educators' specific roles (e.g., what subject and grade level they teach) as well as what the audience already knows and does to grow understanding. Connect why, what, and how to support reasoning.
Be Specific About Harms	Build awareness of how AI can cause various harms through an educational system. Examples include: • Intensifying cyberbullying • Giving erroneous information • Making decisions that limit students' opportunity • Falsely or unfairly flagging detections of cheating or use of AI content
Human Agency and Risk Management	Highlight human agency and oversight in technical processes, help teachers understand their role in risk management, and when and how to engage individuals in other roles.

System Considerations

Fit AI into an Educational Vision	Prioritize AI applications that align with an educational vision and strategy for AI, as well as specifics in the classroom setting. Be wary of chasing trends and be strategic about implementation.
Build on Prior Initiatives	Build upon prior efforts to develop capacity in digital citizenship, media literacy, data science, and computer science.
Engage Relationships	It is too much to ask teachers to shoulder the burden of responsible AI alone. Design a system that engages the educational institutions' relationships to researchers, policy makers, and vendors, so that all support responsible AI together.

People Considerations

Meet Teachers Where They Are	Teachers may have varying attitudes toward and comfort with AI. A district's AI literacy plan for educators should take into consideration a range of participants, including differences in grade level, subject matter, and student population.
Overcome Unequal Resources	Teachers may have differential access to information resources, as well as time and support for developing AI literacy. Plan to overcome preexisting barriers.
Emphasize a Human Agency	Emphasize that humans are guiding, overseeing, and verifying the outputs generated by AI – that humans are in the loop – and are ultimately accountable for how AI they have chosen to use impacts students. However, keep in mind the limits of what teachers (or any person) can reasonably do, given their existing workload.

Module 9: Updating Policies and Advocating for Responsible Use

As AI becomes more prevalent, educational leaders should revisit their existing policies and any related guidance in place regarding the technological tools and platforms used by teachers, administrators and students and consider how to integrate AI and other emerging technologies. The Department urges educational leaders to prioritize developing and updating clear policies that define expectations about the topics covered in this toolkit and to be advocates for responsible AI use in interactions with the local community and with the developers who supply technology used in their schools.

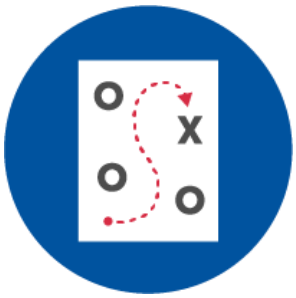
What Is a Responsible Use of Technology Policy?

A **Responsible Use of Technology Policy** (RUP) provides guidance and supporting information to ensure responsible uses of technology systems and tools in schools. The following guidance was created with entire school ecosystems in mind, including staff, students, parents and caregivers, and learning community members. Most schools and districts already have policies governing use of technology. While often referred to as "acceptable use of technology," districts are also using the term "responsible use of technology" to emphasize user agency and voice in using emerging technologies ethically (Ruiz et al., 2024). User agency and voice is particularly important with AI because many tools are entering educational settings from everyday use by individuals but can impact the whole community. A RUP gives users choice but holds them accountable to community standards.

Most schools and districts already have acceptable or responsible use of technology policies in place. To account for the rapid changes in AI and other emerging technologies, the Department recommends that district leaders update their policies.

How to Refine and Implement a RUP

Below, we outline key steps educational leaders can take to develop and maintain responsible use guidelines.

Establish a baseline 	1. Review key components that constitute responsible use of AI: o **Existing** components of the district's RUP. o **Current** Federal, state, and local privacy and civil rights regulations impacting the development of responsible use plans. o **New** opportunities and risks specific to AI (as discussed in other modules), including how peer districts have addressed responsible use. 2. Answer questions from the learning community including:

	○ How is AI being used now? What new uses are likely, whether arriving by student choice, through integration into existing platforms, or via newly acquired tools? ○ What appropriate uses of edtech and AI should continue? (Consider perspectives of students, educators, families, administrators, and others.) ○ What are inappropriate uses of AI that should be prohibited? (Consider multiple perspectives.) ○ What are high-risk uses of AI where additional risk management practices should be required? ○ In what ways can students and educators have agency and choice, while avoiding downsides and risks to the broader community? 3. Collect a set of scenarios to test the proposed policy. A starter set is included in Appendix 3 regarding cyberbullying, accessibility, and surveillance.
Design or update the RUP 	4. Incorporate key elements for a successful responsible use plan: ○ Context for understanding the policy, including the district's vision, principles, values, and understanding of rights and responsibilities. ○ Definitions and/or expectations around AI literacy, including understanding of how technology use can impact well-being. ○ Specific responsibilities and expectations regarding academic behavior (for both educators and students) regarding, for example, how to cite sources, how to represent one's own work, how to respect copyright and avoid plagiarism, and protecting privacy and security. ○ Allowable or prohibited uses of AI, such as allowing the use of AI to generate initial ideas and prohibiting cyberbullying. ○ High-risk uses of AI, such as the use of AI to make decisions which may impact the rights or safety of students and staff. ○ Consequences for misuse, which specifies how the educational institution will monitor use and respond to misuse. ○ Recognition that students and staff with disabilities may have unique needs and might require something in addition to what the RUP provides, and that their needs will be considered consistent with the IDEA, ADA, and Section 504. 5. Include additional guiding language on key topics of responsible AI such as:

	○ Safety, including data privacy and security ○ Accountability ○ Transparency ○ Fairness (and avoiding bias) ○ Efficacy 6. Involve multiple roles in defining and reviewing the policy, and clearly identify how students, educators, administrators, families, and broader community members participate in enacting and supporting the policy.
Implement the RUP and periodically improve it 	7. Disseminate the policy, train community members in its use, and actively follow it. 8. Create feedback loops with educators and students (and the task force recommended in Module 7) to address new or challenging situations as they arise. 9. Establish the cadence in which the learning community will review and update the policies based on use, technological updates, and feedback. 10. Monitor implementation and follow a process for reporting, investigating, and resolving violations of the policy. Collect data on student use and any discipline that may be related to AI to ensure that the use of AI does not result in disproportionate discipline of certain student groups.

Advocating for Responsible Use with Technology Partners

RUPs provide guidance for educators and students to select the appropriate tools for educational tasks, and to use them ethically. RUPs can also be the basis for important discussions with technology partners—the vendors or suppliers of the technologies in use in the district. Technology partners can make it easier for students, educators, and other community members to engage in acceptable use (for example, by making it easier for students to disclose how they used AI in completing an assignment) and prevent prohibited uses.

The Department has produced a guide called Designing for Education with Artificial Intelligence: An Essential Guide for Developers that complements the guidance here. Both documents have foundations in the topics covered in Modules 1-4 of this toolkit, for example. Organizations such as the Software Information and Industry Association have released Principles for the Future of AI in Education, and EdSAFE AI released its SAFE benchmarks. These documents can help educational leaders advocate for responsible AI development and deployment as well as responsible use.

Indeed, trust across educational institutions and technology partners benefits all:

- Both educational institutions and technology partners want meaningful and lasting use of AI to improve education, and they can work together to prioritize a clear vision and avoid distraction by the latest (but unrelated) advances.
- Both educational institutions and technology partners have roles in establishing the safety of AI for educational use, and neither can do all the work on their own.
- Both educational institutions and technology partners have an incentive to address access, use, and design divides so that the full diversity of student populations can benefit.
- Both educational institutions and technology partners strengthen their work by clear communication and feedback loops that can lead to further mitigating risks and making product improvements.

This toolkit recommends that educational leaders use their RUP and other strategic AI documents to advocate with their technology partners for improvements that support the district's vision and strategy for using AI.

Put it into practice! How will your district apply what you have learned in situations involving accessibility, cyberbullying, and surveillance?

See Appendix 3 for scenarios to review and share with your team.

Module 10: Ambitious Action Plans

In this concluding module, the Department calls upon educational leaders to build an organization-wide action plan for implementing AI in their educational community. Educational leaders will be continuously called upon to vet new ways of using AI and to manage anticipated and emergent risks. Doing so responsibly will require strengthening capacities across many roles and responsibilities.

In building a plan, educational leaders should anticipate three ways that AI will enter their community:

1. **Procurement:** Some uses of AI will be formally procured and implemented, as has been the case for edtech for decades.
2. **Integration:** AI will be integrated into existing educational products and services that are already in use.
3. **Diffusion:** AI will spread widely in society and enter schools in the hands of students and educators who bring everyday tools into learning settings.

An organizational plan should address a varied and changing AI landscape and include ongoing reviews to account for AI integration and diffusion. Because most educational institutions have established ways of working with vendors, AI may enter via formal procurement as a separate product less frequently. Based on experience, the Department anticipates that AI will enter schools more frequently through diffusion and integration. This means that teachers, students, and parents will likely use AI-enabled tools that circumvent the traditional approval processes. Increasing awareness about the opportunities and challenges of AI-enabled tools will require the evaluation of these tools by not just IT and emerging technology experts, but also by the broader school community.

Why Build an Organization-Wide Plan for AI?

As discussed throughout this toolkit, the growing presence of AI in education requires addressing longstanding concerns such as data security and privacy as well as newer technical issues such as algorithmic discrimination. AI-specific challenges also include managing new problematic student behaviors such as cyberbullying and academic dishonesty. A look at the history leads to a clear lesson: Educational leaders will need to have a comprehensive plan for responsible AI integration, and the time to start building that plan is now.

What Should be Included in an Organization-Wide Plan?

In its Executive Order on AI, the administration wrote that "Harnessing AI for good and realizing its myriad benefits requires mitigating its substantial risks. This endeavor demands a society-wide effort that includes government, the private sector, academia, and civil society." In an educational organization such as a school district, a responsible AI plan will convey how the district intends to coordinate its people, resources, and activities to realize strategic educational benefits while mitigating the most consequential risks. While the scope of

responsible AI plans is still evolving, some educational organizations have begun to create planning tools. Some examples include:

- The GenAI Readiness checklist, authored by the Council of Great City Schools, CoSN and Amazon Web Services, covers readiness in the following areas: Executive Leadership, Operational, Data, Technical, Security, and Legal/Risk Management.
- Michigan's Virtual Learning and Research Institute produced an AI Integration Framework. This includes many of the categories listed for the checklist (above) and adds areas such as Curriculum and Instruction, Assessment, and Community Outreach.
- TeachAI's Policy Ideas cover five areas – Leadership, Governance, Capacity, AI Literacy, and Supporting Innovation.
- The EDSAFE AI Framework emphasizes working cohesively to advance Safety, Accountability, Fairness and Transparency, and Efficacy.

Individual educators, students, and families often look to school and district leaders to ascertain which AI tools and resources are safe and appropriate to use, and many districts already maintain a list, often on a website, of technologies they have reviewed for use in classroom instruction or other educational activities. Revising and updating these lists is an important starting point for educational leaders. As emphasized throughout this toolkit, educational leaders should evaluate existing resources and build from there.

Beyond initial evaluation, tools like the Michigan Virtual AI Integration Framework recommend a development progression: investigating, implementing, and innovating. This toolkit is aligned with an early developmental stage, in the form of questions for educational leaders who are guiding their communities through an exploratory phase.

Below, we suggest potential steps for an iterative developmental process. To build a comprehensive roadmap for their organization, educational leaders should review all the modules in this toolkit as well as additional resources from organizations they work with and trust.

1. **How does AI impact data privacy and security?** In an initial phase, educational leaders should review their existing data privacy and security policies, as well as guidance and sample policies from their states and from nonprofit organizations before developing new policies. We note three key directions for development:
    a. Guiding educators, students, and community members about the privacy and data security risks that arise in technologies that enter the community, especially through diffusion.
    b. Addressing the increasing breadth of private data that may be collected as edtech becomes more multimodal.
    c. Clarifying the increasing role of biometrics in access and monitoring in educational settings.
2. **How could AI enhance or undermine digital equity and civil rights?** In an initial phase, educational leaders should map the access, use, and design issues (see discussion in

Module 2) already in play in their community. Important topics for development include:

   a. Investigating how algorithms in AI and emerging technologies can lead to discrimination as platforms increasingly use them to suggest resources for students and make recommendations about student placements and evaluations to educators, raising concerns about biased surveillance and inequitable outcomes.
   b. Understanding that teachers are becoming heavily reliant on school-sanctioned AI content detection tools, and there is an increase in student discipline as a result. These AI detection tools have also been found to be ineffective.
   c. Considering how the school will monitor AI usage for discrimination, and how it can be reported, investigated, and remedied when it happens.
   d. Pursuing greater inclusion and use of AI for students with disabilities, multilingual learners, and for other student strengths and needs, especially as AI-enabled technologies could provide students with more ways to effectively engage in learning environments.
   e. Addressing differential access to AI by students and their communities, especially in light of pricing models that may work for some families and not others.

3. **What evidence supports the effective use of AI in education?** In an initial phase, educational leaders should review the available evidence on AI to support their vision and approaches. Three key directions for growth include:
   a. Developing their community's healthy skepticism of industry-based rationales for why and how to use AI in education, which may have emotional appeal but little basis in research.
   b. Strengthening the degree to which evaluation of the rationale for a proposed AI use is based on research, or similarly, on trustworthy summaries of research such as in the What Works Clearinghouse Practice Guides.
   c. Continuing to require high-quality evidence from vendors for the efficacy of their products (U.S. Department of Education, 2024), such as independent research that would meet the Department's standards for Strong or Moderate evidence, if available, including whether any integrated AI features add value and have addressed risks (such as bias). Districts can require evidence from vendors that includes:
      i. Third party certifications
      ii. The creation of a data collection and analysis plan to check outcomes status before renewing a contract
      iii. A logic model and annotated bibliography to understand the research basis
4. **Which opportunities and risks should an educational leadership team prioritize?** Educational leaders should focus their community on opportunities that matter most to their educational vision and approach, and avoid distraction by alluring but misaligned and untested possibilities. Risks are addressed in Module 4 and below. Educational leaders should understand that risks may appear in technologies that enter through any

of the routes (procurement, integration, diffusion), and thus risk management should be comprehensive rather than just a component of formal adoption of technology.

How can educational leaders define an organization-wide plan?

Few educational institutions become "ready" for AI overnight, and most will strengthen their plans iteratively over years. Innovative educational leaders guide their communities through developmental growth in capabilities and structures over time. The Department suggests using the structure of the NIST AI Risk Management Framework (see Module 4) to guide this ongoing process, but with a modification to transform into both an opportunity vetting and a risk management framework. We suggest expanding the four elements of the NIST AI Risk Management framework as follows:

- **Govern:**
  - Develop a culture that assesses opportunities that align with the core of the district's vision and strategy and avoids distraction.
  - Promote a culture where people in all roles in the district have training about risks and contribute to managing risks.
- **Map:**
  - Recognize how opportunities are specific to the varied contexts in the district, such as grade levels, subject areas, and other strengths and needs of groups of students and district community members.
  - Identify and document use, access, and design divides that have impacted equal opportunity to learn in the past, and how use of AI-enabled technologies might address these divides.
  - Anticipate both existing risks (e.g., that an 'adaptive' product necessarily collects information about each student) and a broader set of potential risks for each type or category of use of AI, including risks that go beyond privacy and data security (e.g., bias and algorithmic discrimination).
- **Measure:**
  - Establish relationships with developers of AI-enabled tools that specify and measure both the evidentiary strength of an opportunity and the quality of risk mitigation, starting by advocating for more transparency and accountability from vendors and suppliers.
  - Continue to strengthen leadership within the district with regard to gathering and using evidence (and research) to measure efficacy, as well as for whom and under what conditions an AI-enabled system mitigates risks, so that opportunities can be realized.
- **Manage:**
  - Strengthen capacity to manage implementation of new uses of technology so that benefits to students, educators, and the community are realized equitably.
  - Increase opportunities for professional learning for educators, including their capacity to manage both anticipated and emergent risks, across AI-enabled technologies that arrive in the hands of students and educators.

We give an example below of how an educational leader might use these four elements in a first round of iterative development, understanding that educational leaders will adapt the elements and goals to meet the needs of their districts.

In this scenario, an educational leader might decide to pursue an initial vision such as "every teacher will be better able to specify uses of AI that are supportive of their academic goals, recognize risks of using AI in their academic context, and know how to work with our task force to ensure student safety throughout the school year." With this vision in mind, the educational leader might propose key actions in their first-year plan, such as:

- **Govern:** Our leadership team will establish an AI task force on the academic use of AI (see Module 7) that will be charged with supporting classroom educators in making decisions about how to protect student safety as they explore responsible use of AI, and with reporting to leadership on both the opportunities and challenges that emerge.

- **Map:** Our IT leadership will identify the 12 most common AI tools that educators and students are using for academic content and how uses vary in our district (for example, in different grade levels, in different schools, or for different content). Given these uses and our context, they will also start to identify important risks and communicate about the risks to district leaders, building leaders, and educators.

- **Measure:** Our climate team will upgrade their survey to measure how GenAI uses may be positively or adversely impacting school climate.

- **Measure:** Our IT team will engage vendors of our most used platforms or systems to understand whether they are incorporating AI, in what ways, and what steps they are taking to ensure transparency and accountability.

- **Manage:** We will deliver AI literacy training to all educators in our district, and the training will prepare them to have discussions with students about AI use on academic tasks. The training will support educators to achieve our initial vision and increase their awareness of risks.

- **Manage:** Our task force will respond to risks that arise as educators pursue academic uses of AI, and will draw district leadership's attention to important, unresolved risks.

Put it into practice! How will your district apply what you have learned about informed adoption?

See Appendix 3 for scenarios to review and share with your team.

Appendix 1: Key Terms and Acronyms

The more granular our knowledge base is on any topic, the richer our collaboration and the tighter our bonds are through shared experiences. Mountaineering and climbing have hundreds of terms for gear, technique, and conditions as does the specific computer science subfield of artificial intelligence. Communication is more effective when all parties have a baseline of such terms, and the following list is intended to help build AI literacy among your teammates.

- **Artificial Intelligence (AI):** The term "artificial intelligence" or "AI" has the meaning set forth in 15 U.S.C. 9401(3): a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine- and human-based inputs to perceive real and virtual environments, abstract such perceptions into models through analysis in an automated manner, and use model inference to formulate options for information or action.
- **Generative Artificial Intelligence (GenAI):** The term "GenAI" means the class of AI models that emulate the structure and characteristics of input data in order to generate derived synthetic content. This can include images, videos, audio, text, and other digital content.
- **Educational Technology (EdTech):** Any technologies designed for and employed by educators that can be used to transform school operations, but more so, teaching and learning for the sake of engagement, post-graduation readiness and success, and to make everywhere, all-the-time learning possible for early learners through PK-12, higher education, and adult education.
- **Large Language Models (LLM):** Large language models form the foundation for GenAI systems and are artificial neural networks. At a very basic level, the LLM detected statistical relationships between how likely a word is to appear following the previous word in their training. As they answer questions or write text, LLMs use the model of the likelihood of a word occurring to predict the next word to generate. LLMs are a type of foundation model, which are pre-trained with deep learning techniques on massive data sets of text documents. Sometimes, companies include data sets of text without the creator's consent.
- **User Interface/User Experience (UI/UX) Design:** The human-centered design processes for creating interfaces for both explicitly digital and physical, analog, or other non-digital applications to yield solutions that are both effective and aesthetically pleasing to the end user.
- **Guardrails:** Provide specific boundaries for the use of AI systems and tools. These include specific examples and use cases and are different from policies which provide a framework for the use of AI.

- **Personally identifiable information (PII):** Personally identifiable information (PII) refers to direct identifiers, such as a student's name or address; indirect identifiers, such as a student's date of birth; and any information that, alone or in combination with other information, is linked or linkable to a specific student that would allow a reasonable person in the school community, who does not have personal knowledge of the relevant circumstances, to identify the student with reasonable certainty.
- **Algorithmic Discrimination:** "Algorithmic discrimination" occurs when automated systems contribute to unjustified different treatment or impacts disfavoring people based on their race, color, ethnicity, sex (including pregnancy, childbirth, and related medical conditions, gender identity, intersex status, and sexual orientation), religion, age, national origin, disability, veteran status, genetic information, or any other classification protected by law. Depending on the specific circumstances, such algorithmic discrimination may violate legal protections. Throughout this framework the term "algorithmic discrimination" takes this meaning (and not a technical understanding of discrimination as distinguishing between items).
- **Sensitive Data:** Data and metadata are sensitive if they pertain to an individual in a sensitive domain (defined below); are generated by technologies used in a sensitive domain; can be used to infer data from a sensitive domain or sensitive data about an individual (such as disability-related data, genomic data, biometric data, behavioral data, geolocation data, data related to interaction with the criminal justice system, relationship history and legal status such as custody and divorce information, and home, work, or school environmental data); or have the reasonable potential to be used in ways that are likely to expose individuals to meaningful harm, such as a loss of privacy or financial harm due to identity theft. Data and metadata generated by or about those who are not yet legal adults is also sensitive, even if not related to a sensitive domain. Such data includes, but is not limited to, numerical, text, image, audio, or video data.

Appendix 2: Building on Foundational Initiatives

Proficient hikers who learn to climb open up exponentially more opportunities for exploration around the world. The same is true for the cross-cutting themes in education we are bringing with us into this new era of AI-enabled edtech.

- **Build on foundational skills from existing initiatives.** Districts and schools likely have systems in place to learn capacity building on a variety of literacies:
    - Web and Media literacy, Mozilla Foundation (https://foundation.mozilla.org/en/initiatives/web-literacy/)
    - Digital Citizenship, Common Sense Media (https://www.commonsense.org/ and https://iste.org/digital-citizenship)
    - Computational thinking (https://iste.org/computational-thinking and https://digitalpromise.org/initiative/computational-thinking/)
    - STEM and STEAM initiatives (https://iste.org/steam-education)
    - 21st Century Skills (https://curriculumredesign.org/frameworks/framework-1-0/)
    - Habits of Mind (https://www.habitsofmindinstitute.org/what-are-habits-of-mind/)
    - Portrait of a Graduate (https://www.battelleforkids.org/ and https://digitalpromise.org/2023/08/21/six-attributes-for-portrait-of-a-powerful-graduate/)

Appendix 3: Reacting and Shaping Use Scenarios

The following real-life examples include possible topics that may arise as you consider how to use AI in your school or district.

Each scenario is accompanied by an example of what an educational leader may do in response to the situation. Educational leaders should consider the guidelines from this toolkit in the context of their specific community and audiences when determining how to address a scenario.

Accessibility

Because digitally enhanced learning design has prioritized accessibility for decades, navigating technology for individuals with disabilities is more manageable than ever before. But GenAI's influence upon digital accessibility is multivariate. Emerging technology may provide focused assistance to assist students with disabilities, such as by employing natural language processing and computer vision. But in other instances, technology may not be designed to be accessible to students who learn differently or who have an IEP or 504 plan. Though this work is still emerging and does come with inherent risks, administrators can expect more progress and innovation related to digital accessibility in the future.

Digital Tools Empowering Real-World Accessibility

Your director of special education forwards you an email from an edtech solution provider that claims to support students with vision impairments in a new way. The solution employs computer vision via the user's smartphone camera to help with navigating physical environments as well as interpreting written signage and other materials, giving students more freedom and agency during the school day. As an administrator, you recognize the potential accessibility value but also the human resource value: Such a tool could ostensibly free up staff to assist students in other ways.

What are your next steps?

An Educational Leader might: Ask the provider about: (1) what they will use the recorded data for (Module 10) and (2) how will they ensure and protect student (and family) privacy (Module 2). District leaders should also engage the community, particularly paraprofessionals and caregivers of students who participate in special education programs, to understand their concerns, recommendations, and needs. This can be done through a listening session (Module 5) or by developing a task force (Module 7).

Assessment Accommodations via Emerging Technologies

State testing is approaching, and a popular vendor already trusted with prior solutions has a new suite of GenAI-enabled tools that the vendor claims are aligned with the ESEA/ESSA provisions and provides assessment accommodations tailored to individual students' IEP requirements. With a high percentage of students who have IEPs, this seems like a valid opportunity to ensure students with disabilities can access the assessment.

What are your next steps?

An Educational Leader might: Ask the vendor to provide specific ESEA-aligned documentation of the tools' basis in research and evidence (Module 4) and to provide more details about what they plan to use the recorded data for (Module 10). The leader could also collaborate with special education educators to understand how the tools could be helpful to students with IEPs, while considering what risks should be mitigated for the populations they serve (Module 3).

Cyberbullying

Cyberbullying is an ongoing, challenging issue which has inflicted extreme harm upon students, even resulting in death. As emerging technologies with autonomous agents surface in the GenAI era, cyberbullying may become more volatile. It is the responsibility of educational leaders to maintain awareness and reach out to their community, including task force members, to understand how to address these issues as they come up.

Deep Fake Images from GenAI

A teacher brings to your attention that a student has been targeted by a cyberbullying effort utilizing deep fake images from a GenAI tool and that her parents are on their way to pick her up. You do not know where the images came from, who made them, or how they were shared. Before the first student has made it down to the office, another student is targeted in the same way. And all that is known is that the images are being circulated by other students on their personal devices.

What are your next steps?

An Educational Leader might: Prioritize addressing harms to the students to assure their safety and well-being. Then, proactive communication with students, families, and the greater district community will be important, for example, holding listening sessions with impacted students and families (Module 5). They should work with their IT team to investigate the source of the attack and consider enhancements to the school's data privacy and security policies (Module 2). They can also train and support the school community about responsible digital behavior, including steps to prevent, report, and address cyberbullying and related attacks (Module 3 and Module 8).

Threatening Messages from GenAI

Your high school's administrative assistant comes to school in the morning and, while reviewing the general inbox for parent emails and regular business, finds a threatening email with explicit details about a threat to cause harm to students at the school. It isn't obvious who wrote and sent the threat because they apparently used a chatbot to write the language and multiple measures to cover their IP address and physical location. However, the administrative team is unsure about the extent of safety concerns.

What are your next steps?

An Educational Leader might: Prioritize student safety, privacy, and civil rights (Module 2) by taking immediate action to address risks while ensuring transparent communication with the community (Module 10). Leaders should collaborate with authorities, according to their emergency response plan and in accordance with applicable law, to fully evaluate the threat and keep families informed. They should explore if the email is covered by FERPA, and, if so, if PII from the email will be disclosed, with consideration to FERPA's general consent requirement and exceptions. If safety concerns persist and cannot be fully mitigated, erring on the side of caution may be necessary while respecting students' right to a safe learning environment. Additional information on responding to these types of incidents can be found in the recent CDT report.

Impersonation Campaigns

A group of students receive convincing, but inaccurate, progress reports with personalized commentary allegedly from their school counselor saying they are at risk for failing in certain classes, causing a slurry of visits to the counseling center and inbound calls from parents who were also copied on the emails. In another attack, several students are impersonated by another student using chatbots and other means to create convincing social media posts directed at fellow classmates to obscure the source, convincing the victims that these posts came from real classmates.

What are your next steps?

An Educational Leader might: Assess the risks of these attacks to student well-being and privacy (Module 3) while leveraging evidence-based cybersecurity measures. A leader may also view this as an opportunity to enhance AI literacy programming (Module 8), educating the school community on recognizing AI-driven misinformation and fostering a culture of responsible tech use. Listening sessions (Module 5) could be another way to empower community members to understand and work to address these issues.

Informed Adoption

Some educators have access to the latest technology, resources, and support for sophisticated digital learning experiences. Others get by with the most affordable and readily available edtech solutions. Some even cobble together their own solutions for data tracking via online spreadsheets and forms, which may not be reliable or secure.

The GenAI paradigm shift may cause further separation between such educational leaders' approaches, but it may help others catch up by supplanting knowledge gaps with increasingly capable intelligent assistants. It is incumbent upon administrators to improve their adoption approaches as solution providers update their product roadmaps and marketing campaigns. As with innovations in adventure equipment, astute educational leaders will be able to discern which claims will help them accomplish their vision and mission and which are just smoke and mirrors.

The "Wow" Factor

Using edtech adoption resources available from trusted organizations, your edtech evaluation team has sifted through some of the latest solutions that fit your school's needs and decided on a clear frontrunner. This tool has GenAI components incorporated into its core functionality. The vendor's demonstrations are truly impressive in efficiency and reliability. However, the vendor cannot answer some of your questions about transparency for intellectual property or "trade secret" reasoning.

What are your next steps?

An Educational Leader might: Continue to advocate for clarity from the vendor (Module 10), which is critical for making an informed purchasing decision. If the vendor cannot provide satisfactory answers regarding intellectual property, that may indicate potential risks that outweigh the tool's benefits. With many edtech solutions available, leaders can feel empowered to explore alternative options that align with both educational goals and transparency standards (Module 10).

The New Kid on the Block

You're walking the exhibit hall at a national edtech conference and are drawn into a booth for a new product that clearly uses AI based upon their signage. A rep shows you a few features and mentions machine learning and natural language processing which draws you further into the demonstration because your team has been complaining about not wanting to lag behind in innovation. However, the team on the floor cannot answer simple questions about the underlying AI model or the training data that were used, and thus cannot refute any potential ethical biases that the system could introduce into your school or district.

What are your next steps?

An Educational Leader might: Remain cautious when edtech vendors cannot answer fundamental questions about their AI models or the training data used, especially given the risk of introducing ethical biases (Module 3). Leaders should prioritize products that align with their school's ethical standards (Module 3) and community priorities (Module 5) and feel confident exploring other options if a vendor cannot provide requested information (Module 10).

"...Now with GenAI!"

Your colleagues are excited about advertised GenAI updates coming to the student information system you've been using for some years, because it will allegedly make certain tasks, such as entering grades and preparing end-of-term reports, more efficient and accurate. However, the vendor is offering this updated package at a premium price, increasing the per-student rate by several dollars.

What are your next steps?

An Educational Leader might: Ask for a demonstration and work with their district/school community to carefully evaluate whether the efficiency gains of the tool justify the increased cost. The team should also ask the vendor for clear evidence of benefits and what safeguards are in place to prevent bias (Module 1, Module 3, and Module 10).

Cost-Effectiveness

You are a seasoned administrator, leading a city-based LEA that is relatively underfunded compared to neighboring suburban districts. Staff turnover is above the state average and your team is considering a new counseling solution, which employs a popular LLM with a custom chatbot interface, that is reportedly attuned to the needs of students needing school counseling across a variety of domains, both academic and social-emotional. Early reports and endorsements praise the developer for its innovation. The tool in question is considerably cheaper than hiring even one more counselor.

What are your next steps?

An Educational Leader might: Evaluate the LLM-based chatbot solution by assessing its civil rights implications, ensuring equitable access and privacy protections, while managing risks such as bias and data security (Module 3, Module 10). Gather feedback from the school community, for example through listening sessions (Module 5), and prioritize building AI literacy among staff to ensure the tool's responsible and effective use (Module 8). Above all, the leader and team would keep in mind that while such tools may be able to enhance or supplement the work of human counselors, risks and biases need to be addressed and AI tools should never replace human team members.

It's About Assessment

Educators are concerned about students using GenAI to cheat on formative assessment formats, including short answer questions and essays. Despite efforts to prevent student cheating, there is an ongoing debate about how to fully verify that a student's work is truly their own. On the other hand, educators well versed in modern instructional methods such as project-based and real-world learning are seeing the emergence of GenAI in edtech as an opportunity to redesign common assessment strategies across content areas and grade levels. Perhaps their standards-aligned approaches, with rubrics for measuring student performance and growth, could prove exemplary for updating assessment practices for teachers who have been employing more traditional approaches.

What are your next steps?

An Educational Leader might: Ensure that assessments using GenAI include human safeguards to prevent biased evaluations, particularly for students from historically marginalized groups (Module 3). Encouraging more authentic, qualitative assessments can promote equitable and meaningful measures of student growth. They can provide optional AI literacy workshops for educators and families to support fair and informed assessment practices (Module 8).

Hello, Neighbor!

You hear about a neighboring district that has taken an aggressive approach to exploring and implementing GenAI in its learning experience design. So have your colleagues, as they regularly talk about the innovative opportunities on the other side of town. Yet the culture of your learning community has historically been more conservative and risk averse.

What are your next steps?

An Educational Leader might: Be transparent about their slow and purposeful approach to integrating GenAI, prioritizing solid research and evidence (Module 4) and aligning with the unique needs and values of their community. By engaging the students, families, and educators, they can listen to community concerns and explore the potential of GenAI while fostering trust alongside innovation (Modules 5-7).

Monitoring and Surveillance

Sophisticated solutions have existed for monitoring and surveillance in our schools for years, and AI is already capable of making inferences based on audio and visual data to highlight and describe events involving people, accelerating privacy risks. While educational leaders hope to use such early warning systems to support educators and students, this segment of edtech has potential for bias and misuse.

Students' civil rights, privacy, and well-being are potentially vulnerable, so students need to be able to trust the policies their educators and administrators implement and enforce surrounding online behaviors at school. An educational leader engaging students based on the automated analysis of digital artifacts needs to be weighed against documented evidence of bias in AI training data sets that disproportionately single out students from historically excluded cohorts based on race, ethnicity or nationality, or socioeconomic status.

Educators must always consider the margins for error with AI in monitoring and surveillance scenarios.

Keeping an Eye on Them

One of your assistant principals alerts you that your content moderation solution has reported some suspicious online behavior from "a usual suspect" and that she will be keeping an eye on this particular student. Though well intended, you have also heard about this issue at an education conference and are having second thoughts about that software's value based on ethics alone.

What are your next steps?

An Educational Leader might: Work with a team including the student's teachers to better understand the context of the accusation (Module 7). Additionally, the educational leader can reach out to the content moderation solution provider to better understand their system, including how it flags and what it counts as suspicious activity, as well as how they recommend educational leaders interpret the reports (Module 2 and Module 3). The educational leader can then speak to the student and their family to hear their perspective and use this information to develop guidance and guidelines for how the content moderation solution can be leveraged in the broader student support and well-being ecosystem (Module 9).

Digital Profiling

Your Chief Technology Officer has been in conversation with an edtech sales representative with an innovative solution that uses AI to interpret the physical presence data from multiple sources including security cameras, Wi-Fi access points, and student ID badge terminals. The solution can accurately track a student's movements and recognize patterns aligned with staff documentation and indicate which students are most likely to misbehave.

What are your next steps?

An Educational Leader might: First, educational leaders must check the laws in their state and city to better understand the legality of this type of surveillance within a school setting (Module 2). They must also consider student privacy and data rights, and whether or not they have permission to use these systems for these purposes (Module 2). Importantly, educational leaders should work to build a shared understanding of how this use of emerging technology works within their context, bringing in community members including students (Module 8) and the technology task force (Module 7).

De-anonymizing Anonymous Posts

A frustrated high school student anonymously posts online a verbose criticism of school policy or even your leadership for being unfair in some manner. While pointed, the manifesto isn't particularly dangerous or threatening. The teenager is using a social media account to share their heated perspective that may have resulted from stress, an isolated incident, or another trigger. However, a recent monitoring tool that has been updated to employ an LLM "de-anonymizes" the student's writing based on other posted language and shares that student's name with you and their advisory teacher.

What are your next steps?

An Educational Leader might: An important first step is to help the district community learn about and better understand the increasing availability of tools that allow for triangulation of data and what that means for them in various school and personal settings (Module 8). Many students and community members do not realize that the large amount of data they make available to companies can be connected easily to identify them even when they are under the impression that they are acting anonymously. This situation also requires the educational leader to work with the appropriate school team to address this situation with the student and their family directly (Module 7).

References

The Association for Supervision and Curriculum Development, The International Society for Technology in Education, The National Association of Secondary School Principals, The National Association of Elementary School Principals; The American Association of School Administrators. (2023, July). Bringing AI to school: Tips for school leaders. https://cms-live-media.iste.org/Bringing_AI_to_School-2023_07.pdf

California Department of Education. (2024, June 12). Learning with AI, Learning about AI. https://www.cde.ca.gov/pd/ca/cs/aiincalifornia.asp

CAST (2024). Universal Design for Learning Guidelines version 3.0. Retrieved from https://udlguidelines.cast.org

Center for Democracy and Technology. (2023, December 18). *The shortcomings of generative AI detection: How schools should approach declining teacher trust in students.* Center for Democracy and Technology. https://cdt.org/insights/the-shortcomings-of-generative-ai-detection-how-schools-should-approach-declining-teacher-trust-in-students/

Chine, D.R. et al. (2022). Educational Equity Through Combined Human-AI Personalization: A Propensity Matching Evaluation. *Rodrigo, M.M., Matsuda, N., Cristea, A.I., Dimitrova, V. (eds) Artificial Intelligence in Education*. AIED 2022. Lecture Notes in Computer Science, vol 13355. Springer, Cham. https://doi.org/10.1007/978-3-031-11644-5_30

Coburn, C. E., & Penuel, W. R. (2016). Research–practice partnerships in education: Outcomes, dynamics, and open questions. *Educational researcher*, 45(1), 48-54.

Cohen, D. K., & Ball, D. L. (1999). *Instruction, capacity, and improvement*. CPRE Research Reports. Retrieved from https://repository.upenn.edu/handle/20.500.14332/8434

Consortium for School Networking, Council of the Great City Schools. (2024b, April). CoSN/CGCS K-12 Gen AI Maturity Tool. https://www.cosn.org/wp-content/uploads/2024/04/Cosn_CGCS-K-12-Gen-AI-Maturity-Tool-Conference.pdf

Consortium for School Networking. (2023, October). Readiness Checklist. https://www.cgcs.org/genaichecklist

Digital Promise. (2024, June 19). District Resources. *Product Certifications*. https://productcertifications.digitalpromise.org/resources/district-resources/

Digital Promise. (2024, March 22). A process to help run successful Ed-Tech pilots. *Edtech Pilot Framework*. https://edtech.digitalpromise.org/

Druga, S., Yip, J., Preston, M., & Dillon, D. (2021). The 4As: Ask, adapt, author, analyze - AI literacy framework for families. In *Algorithmic Rights and Protections for Children*. https://doi.org/10.1162/ba67f642.646d0673

The Edtech Equity Project. (2020, June). *School Procurement Guide: Buying Edtech Products with Racial Equity in Mind*. The Edtech Equity Project. https://www.edtechequity.org/work/school-procurement-guide

TNTP. (2018). Choosing the Opportunity Gap. *The Opportunity Myth.* https://opportunitymyth.tntp.org/choosing-the-opportunity-gap

Elmore, R. (2008). *Improving the Instructional Core*. Harvard University, School of Education. Retrieved from https://www.pps.net/site/handlers/filedownload.ashx?moduleinstanceid=19185&dataid=123381&FileName=Improving_the_Instructional_Core.pdf

EngageAI Bill of Rights Task Force. (2024, May). An AI Bill of Rights for Educators. EngageAI Institute.

Farrell, C. C., Penuel, W. R., Allen, A., Anderson, E. R., Bohannon, A. X., Coburn, C. E., & Brown, S. L. (2022). Learning at the boundaries of research and practice: A framework for understanding research–practice partnerships. *Educational Researcher*, 51(3), 197-208.

Gao, G., Yang, X., & Chi, M. (2024). Get a head start: On-demand pedagogical policy selection in intelligent tutoring. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(11), 12136-12144. https://doi.org/10.1609/aaai.v38i11.29102

Kentucky Department of Education. (2010). *Highly Effective Teaching and Learning in Kentucky Overview Guide.* https://www.education.ky.gov/curriculum/standards/teachtools/Documents/Highly_Effective_Teaching_and_Learning_Overview_Guide.pdf

Kentucky Department of Education. (2024). AI Guidance Brief. https://www.education.ky.gov/districts/tech/Documents/AI%20Guidance%20Brief.pdf

Kulesa, A. C., Croft, M., Brian, R. and Wells, M. K. (2024). Learning Systems: Opportunities and Challenges of Artificial Intelligence-Enhanced Education. Bellwether. Retrieved October 14, 2024, from https://bellwether.org/wp-content/uploads/2024/09/LearningSystems_2_Bellwether_September2024.pdf

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42-78. https://doi.org/10.3102/0034654315581420

Laird, E., Dwyer, M., & Woelfel, K. (2024, September 26). In Deep Trouble: Surfacing Tech-Powered Sexual Harassment in K-12 Schools. Center for Democracy & Technology. https://cdt.org/wp-content/uploads/2024/09/2024-09-26-final-Civic-Tech-Fall-Polling-research-1.pdf

Mills, K., Ruiz, P., Lee, K., Coenraad, M., Fusco, J., Roschelle, J. & Weisgrau, J. (2024, May). AI literacy: A framework to understand, evaluate, and use emerging technology. https://doi.org/10.51388/20.500.12265/218

MIT Media Lab Personal Robots Group and the MIT STEP Lab (2024). The Daily Curriculum for middle school students. https://raise.mit.edu/daily/

National Center on Improving Literacy (2022). The Science of Reading: The Basics. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education, Office of Special Education Programs, National Center on Improving Literacy. https://improvingliteracy.org/brief/science-reading-basics/index.html

National Educational Technology Plan. Office of Educational Technology. (2024, January 24). https://tech.ed.gov/netp/

National Institute of Standards and Technology. (2023). AI Risk Management Framework. *Artificial Intelligence Risk Management Framework* (AI RMF 1.0). https://doi.org/10.6028/nist.ai.100-1

NGSS Lead States. (2013). Next Generation Science Standards: For States, By States. Washington, DC: The National Academies Press.

Office for Human Research Protections (OHRP). (2024, July 10). *The Belmont Report*. Department of Health & Human Services. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html

Office of Management and Budget. (2024, March 28). Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence (Memorandum M-24-10). The White House. https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf

Roschelle, J., Fusco, J., & Ruiz, P. (2024). Review of guidance from seven states on AI in education. Digital Promise. doi.org/10.51388/20.500.12265/204

Ruiz, P., Armstrong, A., Karim, S., Shell, A., Singmaster, H., & Giang, M. (2024) Emerging technology acceptable use policy [Google slides]. Digital Promise. https://bit.ly/DPAUPLanguage

Ruiz, P., & Fusco, J. (2024, March 31). Glossary of Artificial Intelligence Terms for Educators. CIRCLS. https://circls.org/educatorcircls/ai-glossary

Southern Education Foundation. (2024, January). Outcomes Based Contracting. https://obc.southerneducation.org/

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology*, 105(4), 970

U.S. Commerce Secretary Gina Raimondo Announces Expansion of U.S. Ai Safety Institute Leadership Team. National Institute of Standards and Technology. (2024, April 23). https://www.nist.gov/news-events/news/2024/04/us-commerce-secretary-gina-raimondo-announces-expansion-us-ai-safety

U.S. Department of Education, Office of Educational Technology, Designing for Education with Artificial Intelligence: An Essential Guide for Developers, Washington, D.C., 2024.

United States Department of Education. (2023, May 24). *Artificial Intelligence and the future of teaching and learning*. Office of Educational Technology. https://tech.ed.gov/ai-future-of-teaching-and-learning/

United States Government Accountability Office. (2022, October). *Critical Infrastructure Protection: Additional Federal Coordination Is Needed to Enhance K-12 Cybersecurity*. https://www.gao.gov/assets/gao-23-105480.pdf

The United States Government. (n.d.). Blueprint for an AI Bill of Rights Definitions. The White House. https://www.whitehouse.gov/ostp/ai-bill-of-rights/definitions/

The United States Government. (2023, October 30). Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. The White House. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

Weiss, A. (2024, July 31). From Vision to Practice: Part 1. Grounding AI for Schools in Learning Science with Dr. Kristen DiCerbo. Khan Academy. https://blog.khanacademy.org/from-vision-to-practice-part-1-grounding-ai-for-schools-in-learning-science-with-dr-kristen-dicerbo/

What Works Clearinghouse. (n.d.). *Institute of Education Sciences.* U.S. Department of Education. Retrieved from https://ies.ed.gov/ncee/wwc/

World Economic Forum. (2024, April). Shaping the Future of Learning: The Role of AI in Education 4.0. https://www.weforum.org/publications/shaping-the-future-of-learning-the-role-of-ai-in-education-4-0/

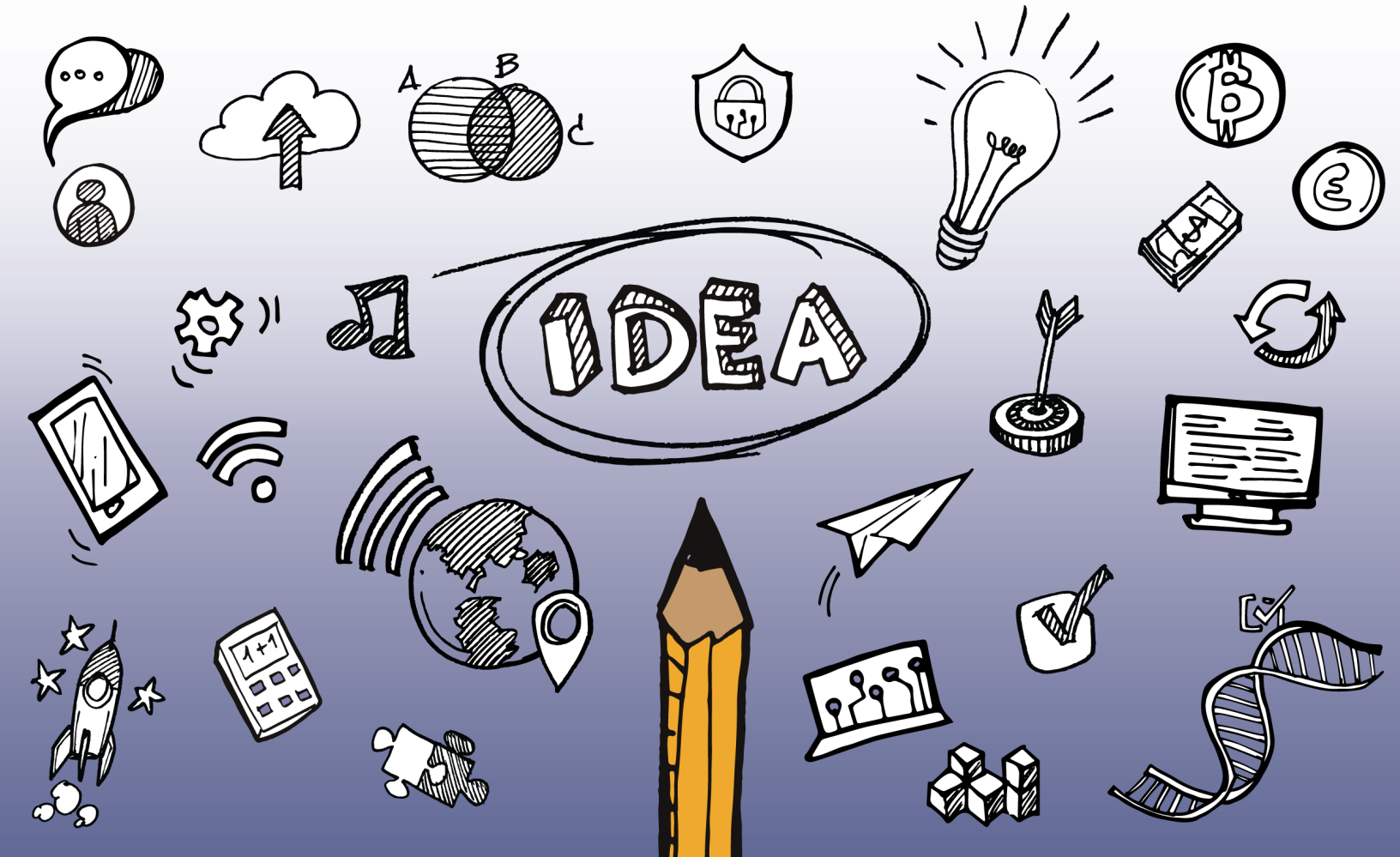
Xu, Z., Wijekumar, K., Ramirez, G., Hu, X. and Irey, R. (2019). The effectiveness of intelligent tutoring systems on K-12 students' reading comprehension: A meta-analysis. *Br J Educ Technol*, 50: 3119-3137. https://doi.org/10.1111/bjet.12758

OFFICE OF
Educational Technology

Navigating Artificial Intelligence in Postsecondary Education:

Building Capacity for the Road Ahead

January 2025



Navigating Artificial Intelligence in Postsecondary Education: Building Capacity for the Road Ahead.

Miguel A. Cardona, Ed.D.

Secretary, U.S. Department of Education

Roberto J. Rodríguez

Assistant Secretary, Office of Planning, Evaluation and Policy Development

January 2025

Examples Are Not Endorsements This document contains examples and resource materials that are provided for the user's convenience. The inclusion of any material is not intended to reflect its importance, nor is it intended to endorse any views expressed or products or services offered. These materials may contain the views and recommendations of various subject matter experts as well as contact addresses, websites, and hypertext links to information created and maintained by other public and private organizations. The opinions expressed in any of these materials do not necessarily reflect the positions or policies of the U.S. Department of Education. The U.S. Department of Education does not control or guarantee the accuracy, relevance, timeliness, or completeness of any information from other sources that are included in these materials. Other than statutory and regulatory requirements included in the document, the contents of this guide do not have the force and effect of law and are not meant to bind the public.

Contracts and Procurement

This document is not intended to provide legal advice or approval of any potential federal contractor's business decision or strategy in relation to any current or future federal procurement and/or contract. Further, this document is not an invitation for bid, request for proposal, or other solicitation.

Licensing and Availability

This guide is in the public domain and available on a U.S. Department of Education website at https://tech.ed.gov.

Requests for alternate format documents such as Braille or large print should be submitted to the Alternate Format Center by calling 1-202-260-0852 or by contacting the 504 coordinator via email at om_eeos@ed.gov.

Notice to Persons with Limited English Proficiency

If you have difficulty understanding English, you may request language assistance services for U.S. Department of Education information that is available to the public. These language assistance services are available free of charge. If you need more information about interpretation or translation services, please call 1-800-USA-LEARN (1-800-872-5327) (TTY: 1-800-437-0833); email us at Ed.Language.Assistance@ed.gov; or write to U.S. Department of Education, Information Resource Center, LBJ Education Building, 400 Maryland Ave., SW, Washington, DC 20202.

How to Cite

Permission to reprint this publication is not necessary. The suggested citation is as follows: U.S. Department of Education, Office of Educational Technology, *Navigating Artificial Intelligence in Postsecondary Education: Building Capacity for the Road Ahead*, Washington, DC, 2024.

Acknowledgements

Project Team

Navigating Artificial Intelligence in Higher Education: Building Capacity for the Road Ahead was developed under the leadership and guidance of Roberto J. Rodríguez, Assistant Secretary for the Office of Planning, Evaluation and Policy Development, Anil Hurkadli, Interim Deputy Director for the Office of Educational Technology, Bernadette Adams, Senior Policy Advisor, Kevin Johnstun, Management and Program Analyst, and Isabella Zachariah, Education Pioneer Fellow, the Office of Educational Technology.

This work was developed with support from Digital Promise under contract (9190019C0076) and led by **Vanessa Peters Hinton** with **Anthony Baker**, **Barbara Means**, and **Jeremy Roschelle**, with assistance from **Rohit Kataria** of Vanderbilt University.

Executive Summary

Artificial intelligence (AI) is transforming many institutional functions at the postsecondary education level, including admissions, enrollment, academic advising, and learning environments. This brief, *Navigating Artificial Intelligence in Postsecondary Education: Building Capacity for the Road Ahead*, aims to support institutional leaders who are engaged in a cross-cutting team (such as an AI task force) that is overseeing the implementation of AI across multiple areas of their institution.

This brief assumes readers have some familiarity with AI's benefits and risks in the context of postsecondary education. We address topics that are relevant to a wide variety of institution types and provide content that is accessible to education leaders who may not have deep experience with or specific academic training in the field of AI. Because we recognize that institutions have diverse financial capabilities, the recommendations also consider both high-resource and low-resource settings to provide scalable solutions for varying budgets.

These core insights are based on the set of recommendations that are outlined in the May 2023 report, *Artificial Intelligence and the Future of Teaching and Learning*. The guidance also aligns with President Biden's Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, focusing on enhancing student learning and institutional effectiveness while ensuring equity, fairness, and the non-discriminatory use of AI.

The U.S. Department of Education (Department) received and considered input from numerous stakeholders across the education landscape in developing its recommendations. By focusing on responsible and ethical AI practices, the brief aims to help institutions prepare learners, administrators, and faculty for changes that will come from AI-driven innovations and leverage AI to improve access and academic success, especially for underserved communities.

Key Recommendations

- **Recommendation 1: Establish transparent policies for how AI is used to support operational activities in postsecondary education settings.** AI policies can cover core areas such as managing admissions, enrollment, and resourcing decisions and help ensure fairness, accountability, and trust by allowing stakeholders to understand how data are used to support decision-making and how decisions are made.
- **Recommendation 2: Create or expand infrastructure to support the innovative application of AI in instruction, student advising and support, and assessment.** Infrastructure should support and encourage AI use that aligns with a shared vision for education and enhances educational quality and student success. Infrastructure should include support for experimental projects and professional development for AI integration by faculty and staff.

- **Recommendation 3: Rigorously test and evaluate AI-driven tools, supports, and services.** Conduct rigorous research and evaluation studies on AI-driven platforms prior to deployment. Use continuous improvement methods to ensure their effectiveness, safety, and alignment with educational goals and student needs.
- **Recommendation 4: Seek collaborative partners for designing and iteratively testing AI models across educational applications.** Forge partnerships with industry, nonprofit organizations, and other postsecondary institutions on AI design and testing, including AI use for enhancing learning experiences, improving institutional processes, and supporting diverse student needs.
- **Recommendation 5: Review, refine, and supplement program offerings in light of the growing impact of AI on future jobs and career opportunities.** Institutions may regularly assess current programs and potentially create new programs to equip both students and workers with the skills necessary for a job market increasingly influenced by AI.

Introduction

AI advancements are likely to have a profound impact on all aspects of society and every sector of the economy in the coming years, with changes already well underway. In this context, postsecondary institutions have a dual role: (1) to strategically leverage AI to help catalyze greater access and success into postsecondary education for all students, especially those from historically underserved groups; and (2) to prepare postsecondary students for the evolving career landscape shaped by AI-driven innovation. This brief, building on the Department's earlier publication, *Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations*, provides guidance for education leaders on implementing AI applications in postsecondary settings. The Department guidance provided here is aligned with the principles and areas of emphasis outlined in the Raise the Bar: College Excellence and Equity and the Raise the Bar: Unlocking Career Success initiatives.

The insights here are intended to support institutional leaders who are engaged in a cross-cutting team (such as an AI task force) that is overseeing the implementation of AI across multiple areas of their institution. We address topics that are relevant to a wide variety of institution types and provide content that is accessible to education leaders who may not have deep experience with or specific academic training in the field of AI. We hope the following analysis and recommendations will help institutions prioritize and collaborate, including with other institutions, as they develop policies concerning the latest wave of AI tools and build enduring capacity that will allow institutions to adapt to rapid and significant technological change in the years to come.

The brief is divided into two major sections. The first provides concrete recommendations, guiding questions for leadership and faculty, and resources to help postsecondary education institutions enhance learning outcomes, improve institutional operations, and prepare students for a future that is increasingly defined by fast-paced technological change. These recommendations are inclusive and adaptive, accommodate varying levels of interest and expertise, and are designed for both high-resource and low-resource institutions to scale solutions. That said, we also recognize that some institutions may face constraints in implementing some of these recommendations. Agencies across the Federal Government, including the Department, have invested in responsible AI initiatives, and the Department encourages other organizations across the ecosystem with resources and expertise related to AI to actively collaborate with historically underserved institutions to identify ways of closing the relevant divides between institutions.

The second section is an addendum presenting evidence-based insights on AI integration. This non-exhaustive literature review examines AI's potential impact on functions such as learning environments, career readiness, admissions, enrollment, student advising, student support, digital infrastructure, and faculty development. These insights informed the final recommendations of the brief, and we hope the addendum will provide institution leaders with information for making evidence-based decisions in implementing those same recommendations.

The brief presents a broad vision of AI technologies and capabilities, based on the recommendations and a literature review, offering evidence-based insights to help institutional

leaders address policy issues and leverage AI in ways that support holistic student development and success. Across both sections of the brief, the Department stresses several key themes:

- The use of AI can support adaptivity and responsiveness to students' various needs, as long as the use of AI tools aligns with a shared vision for the education institution and leverages evidence-based principles.
- Outputs and recommendations from AI tools should be appropriately balanced with human oversight to support insightful and informed decision-making that aligns with educational goals and ethical standards.
- AI platforms should augment, not replace, faculty and administrative expertise in leading operations, providing instruction, and scaffolding student supports.
- To seize the benefits of AI, institutions should test and evaluate AI platforms and build faculty, staff, and student capacity to support their safe, responsible, and non-discriminatory use.

With these themes in mind, this brief explores broad applications of AI in higher education as described in published literature and key stakeholder engagements. The studies included here prioritize peer-reviewed articles and white paper reports of recent AI work, especially those that implemented experimental research designs. Throughout the document, the Department has included insights from roundtable discussions with faculty, staff, researchers, administrators, and representatives of labor unions from an array of institutions as well as from the AI in Postsecondary Education Working Session convened at the White House by the Department in June 2024. The roundtable discussions and the AI Working Session included representatives from large research universities, four-year teaching universities, fully online institutions, community colleges, Historically Black Colleges and Universities, Hispanic Serving Institutions, Tribally Controlled Colleges and Universities, and career and technical education programs.

This guidance is responsive to the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (Oct. 23, 2023), which states (Section 8(d)):

To help ensure the responsible development and deployment of AI in the education sector, the Secretary of Education shall, within 365 days of the date of this order, develop resources, policies, and guidance regarding AI. These resources shall address safe, responsible, and nondiscriminatory uses of AI in education, including the impact AI systems have on vulnerable and underserved communities, and shall be developed in consultation with stakeholders as appropriate.

This brief also supports efforts aligned with the Executive Order's directive regarding retooling career pathways in response to AI advances. Specifically, the Executive Order requires the Secretary of Labor to (Section 6(a)(ii)(B)):

> identify options, including potential legislative measures, to strengthen or develop additional Federal support for workers displaced by AI and, in consultation with the Secretary of Commerce and the Secretary of Education, strengthen and expand education and training opportunities that provide individuals pathways to occupations related to AI.

This brief provides non-regulatory, education-specific guidance that is aligned with federal guidelines and guardrails. Coverage of existing federal guidelines and guardrails is not comprehensive or exhaustive. It is not intended to and does not enable a postsecondary education institution to establish its compliance with regulations. Also, it is not intended to and does not introduce any new requirements. Where examples are given, including links to resources outside of the U.S. Government, they are intended to be illustrative and not to restrict the application of this guide to additional forms of AI as they become available for use in education. We are providing these external links because they contain additional information relevant to the topic(s) discussed in this document or that otherwise may be useful to the reader. We cannot attest to the accuracy of information provided on the cited third-party websites or any other linked third-party site. We are providing these links for reference only; linking to external resources does not constitute an endorsement by the Department. Education leaders can use this guide to increase their understanding of essential federal guidelines and guardrails to guide their work as they create AI applications for educational settings. Readers are also encouraged to engage with other higher education institutions and organizations to refine their approaches toward responsible AI use.

Five Recommendations for Integrating AI in Postsecondary Education

The integration of AI in postsecondary education presents a unique opportunity to enhance student supports and services, improve academic outcomes and learning, foster innovative research and development, and strengthen career preparedness for a global and digital economy. Achieving each of these potential benefits requires careful foresight and diligent planning. We offer the following recommendations to support leaders in their efforts.

Recommendation 1: Establish transparent policies for how AI is used to support operational activities in postsecondary education settings. AI policies can cover core areas such as managing admissions, enrollment, and resourcing decisions and help ensure fairness, accountability, and trust by allowing stakeholders to understand how data are used to support decision-making and how decisions are made.

While AI can improve the quality of admissions and enrollment processes,[1] as well as the quality of resourcing decisions more generally, institutions should have policies for safe and transparent use of AI to avoid perpetuating biases and exacerbating inequitable outcomes for students.

To avoid these unintended negative outcomes, institutions' AI policies should emphasize transparency, accuracy, privacy, and ethical considerations at each stage of the AI platform

[1] For example, Georgia State University leveraged an artificial intelligence as part of a systemic approach to addressing the attrition rates of students who begin the enrollment process, but do not attend during that same semester.

development or adoption processes. The National Institute of Standards and Technology (NIST) provides guidance through its *AI Risk Management Framework* (AI RMF) to help all types of organizations to understand and manage risks associated with AI by focusing on four key components of the AI lifecycle: Govern, Map, Measure, and Manage. The framework balances AI benefits with societal protection and is designed to adapt to rapid advancements and the evolving nature of AI technologies. The AI RMF is a valuable resource, but it is not education specific, so institution leaders should also seek insights from education-focused organizations, engage peer institutions, and openly communicate their experiences to promote shared learning.

Further, the Office of Management and Budget recently released the Advancing the Responsible Acquisition of Artificial Intelligence in Government memorandum (OMB M-24-18) which provides best practices for managing AI risks, particularly when it comes to rights and safety. The memorandum emphasizes strong contractual terms to protect data, ensure safe AI use, and manage its impact on public decision-making. We recommend similar considerations for institution leaders to ensure responsible AI use in education, safeguarding students' rights and privacy.

A detailed analysis of each potential aspect of a robust AI policy is beyond the scope of this brief. However, given the central role of establishing transparency through stakeholder engagement in both establishing and evaluating any AI policy, we offer two helpful examples related to emerging technological capabilities, "stealth assessment" and "continuous monitoring." Both of these examples demonstrate how a lack of transparency can erode trust among stakeholders and undermine the institution's commitment to fairness and inclusivity.

1) In its original meaning, "stealth" was a synonym for "unobtrusive" and was intended to provide authentic and supportive feedback to students while maintaining uninterrupted high student engagement in learning.[2] However, "stealth" can also imply surveillance—that students may be measured invisibly, without their knowledge, and with no direct and obvious benefit to their learning. With regard to the original meaning, students and faculty want to reduce the amount of time taken away from learning for testing and thus may appreciate unobtrusive assessment that can be embedded in learning activities, especially when it's clear how the outputs help faculty and students immediately and directly. Yet, with regard to the second meaning, faculty and students may be rightly concerned if sensitive learning data is shared beyond the classroom without their knowledge, and that could have unforeseeable consequences for the students' performance, well-being, and future opportunities. As with all data collection, clear disclosure of use and affirmative consent is of primary importance. Additionally, clearly communicating the purposes and limits of unobtrusive data collection—as well as involving teachers and students in meaningful data use—can make the difference in trust.
2) In response to campus safety concerns, some institutions have turned to AI-enabled continuous monitoring services. These services tend to leverage algorithms that have been trained to identify certain indicators of suspicious activity and can be deployed to analyze

[2] Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503-524). Information Age Publishers.

real time video feeds and provide "early warnings" connected to a potential security concern. However, in cases where there is not robust input from the campus community on whether or not to roll out the system, it can result in a reputational harm to the institution. These systems often require the collection of sensitive biometric data (face, gait, voice, location). Collecting these forms of data can result in significant harm if, for example, it led to cases where individuals were misidentified leading to an unwarranted interaction with law enforcement.[3,4] Furthermore, because of the multiple concerns related to safety, privacy, and civil liberties, the White House stated in the Blueprint for an AI Bill of Rights: "Continuous surveillance and monitoring should not be used in education, work, housing, or in other contexts where the use of such surveillance technologies is likely to limit rights, opportunities, or access. Whenever possible, you should have access to reporting that confirms your data decisions have been respected and provides an assessment of the potential impact of surveillance technologies on your rights, opportunities, or access." (p. 6) Thus any proposal to use AI in support of campus safety should also include an open and transparent process to assess these concerns.

Both of these examples demonstrate the need for promoting transparency through comprehensive stakeholder engagement to inform what AI systems are utilized, what data they collect, and how the data influence decision-making.

Key Questions:

- How could the Blueprint for an AI Bill of Rights or NIST AI RMF help your institution develop and adopt trustworthy AI systems?
- How can your institution work to promote transparency in its procurement and adoption of AI tools and technologies?
- How can institutions evolve their disclosure and affirmative consent practices around the use of AI to reflect the values of the campus community and uphold civil rights and civil liberties?
- How can your institution review and document the characteristics of trustworthy AI systems in its AI-driven products and services?
- How can your institution take ownership of its use of AI, ensuring that administrators, staff, and students throughout the organization understand the data being used and the tasks being supported?

Resources:

- The Blueprint for an AI Bill of Rights outlines five principles to guide the design, use, and deployment of automated systems, including AI.
- The EDSAFE AI Alliance has developed a comprehensive framework and guidelines for the use of generative AI across various stakeholder groups in the education sector. It is designed

[3] Weinstein, M. (2020). School surveillance: The students' rights implications of artificial intelligence as K-12 school security. *North Carolina Law Review, 98*(2), 438.

[4] Hassanin, N. (2023, August 23). *Law professor explores racial bias implications in facial recognition technology.* University of Calgary News.

to support AI policy labs and other stakeholders in understanding and addressing the opportunities, challenges, and ethical considerations associated with AI deployment in educational contexts.

- The CARE Principles for Indigenous Data Governance support Indigenous communities' self-determination, emphasizing data sovereignty and informed governance, developed through input from diverse experts and stakeholders.
- The Center for Innovation, Design, and Digital Learning published *Inclusive Intelligence: The Impact of AI on Education for All Learners*. The report examines AI's role in education, highlighting its potential while addressing challenges for students with disabilities and diverse learning needs.
- Complete College America's Attainment with AI playbook provides practical ideas for higher education institutions and systems to employ generative AI for student success, completion, and equity.

Recommendation 2: Create or expand infrastructure to support the innovative application of AI in instruction, student advising and support, and assessment. Infrastructure should support and encourage AI use that aligns with a shared vision for education and enhances educational quality and student success. Infrastructure should include support for experimental projects and professional development for AI integration by faculty and staff.

Supporting the safe and fair use of AI in postsecondary education requires both digital infrastructure and human infrastructure. The EDUCAUSE Higher Education AI Readiness Assessment helps institutions determine their preparedness for strategic AI initiatives across both types of infrastructure.[5] With respect to digital infrastructure, robust and secure digital networks are essential to protect sensitive data and maintain privacy. Indeed, comprehensive cybersecurity measures are also crucial to protect against potential threats and disruptions.

To build or customize AI systems, institutions need access to robust computing resources, including cloud computing services,[6] to effectively run AI applications. These resources are essential for processing large datasets, conducting real-time analytics, and integrating AI into academic and administrative functions. Cloud computing can also provide additional processing power, scalability, and storage capabilities that surpass or augment on-site hardware, ensuring that AI models can be built and customized efficiently. It is important to note that training, updating, and leveraging AI models can be energy resource intensive. Institutions should monitor this activity and align their efforts with their broader goals of reducing environmental impacts.

The Department notes there are wide disparities between institutions in terms of access to the infrastructure needed to safely and effectively leverage AI. Consortia and resource centers that

[5] EDUCAUSE. (2024, April). *Higher Education Generative AI Readiness Assessment.*

[6] Cloud computing provides on-demand access to remote servers over the internet, allowing institutions to make use of scalable processing power, vast storage, and advanced networking, making it possible to run AI applications, process large datasets, and perform real-time analytics without the limitations of traditional on-site hardware.

bring together a broad array of diverse institutions (such the National Artificial Intelligence Research Resource described below) are being organized to help to reduce such disparities, but more effort and attention is needed in this area to avoid exacerbating existing digital divides.

Just as the Department is concerned about divides between institutions, institutions should also take care to promote practices that enhance digital equity[7] among students and staff. In many cases this will include ensuring that students and staff have equitable access to internet connectivity, digital devices, and AI platforms that can help enhance their learning and career opportunities.

However, digital infrastructure alone won't lead to equitable AI innovation. Planning to support human infrastructure is also needed to equip students, faculty, and staff with essential AI literacy skills for safe and effective use of the technology.[8] AI literacy involves the knowledge and skills to understand, use, and evaluate AI systems critically, promoting safety and ethics in communities. This includes recognizing and addressing how AI technologies can facilitate discrimination and harassment, such as through the nonconsensual distribution of images that have been altered or generated by AI technologies. Key activities for AI literacy programs include understanding, evaluating, and integrating AI into learning environments, operations, and student support systems. Education leaders are in the best position to decide how to cover these areas based on their specific contexts. Students will undoubtedly enter postsecondary education with vastly different levels of AI literacy. Non-traditional students may experience particular challenges in developing AI literacy skills since they may have delayed enrollment in postsecondary education after high school, making it harder to adapt to rapidly evolving technological concepts. To support students equitably and fairly, AI literacy approaches should be flexible enough to benefit learners entering postsecondary school at different stages of their life and educational journey.

Institutions should provide support for general and discipline-specific professional development for faculty and staff that addresses the use of AI tools in teaching and learning. This can be accomplished through an appropriate office (such as a center for academic innovation, libraries division, or academic technology office) that conducts a needs assessment and develops a comprehensive faculty development plan for AI in teaching and learning. Promising AI tools and applications will vary across different academic disciplines, suggesting that at least part of the professional development be designed for groups of faculty within the same discipline or program.[9]

Faculty should be given the time and opportunity to collaborate with their peers to learn how to implement broad-use AI models in their teaching and research. Faculty, staff, and students should also be provided the opportunity to give input into how AI is used in teaching and learning. These opportunities should provide faculty with time to grapple with substantive questions like how to define appropriate AI use for a given learning objective, how to integrate AI-driven assistive

---

[7] Section 60302(10) of the Infrastructure Investment and Jobs Act defines "digital equity" as "the condition in which individuals and communities have the information technology capacity that is needed for full participation in the society and economy of the United States."

[8] Office of Education Technology. (2024, April). *AI Literacy 101: An Introduction to Artificial Intelligence*. U.S. Department of Education.

[9] Lee, L-K., Cheung, S. K. S., & Kwok, L-F. (2020). Learning analytics: Current trends and innovative practices. *Journal of Computers in Education, 7*(1), 1–6.

technologies that support students with disabilities or English language learners, how to evaluate the results of an AI detection tool, and how to build strong relationships with students that can foster trust and integrity as everyone adapts to new norms.

One example of a successful culture of innovation can be seen in Carnegie Mellon University's Generative AI Teaching as Research (GAITAR) programs, which are designed to build instructor capacity for adapting, innovating, and conducting scholarship on the use of generative AI in teaching and learning. Participants of the GAITAR program attend workshops and consultations to develop testable research questions, design teaching innovations and interventions, identify data sources, and plan classroom research studies to rigorously evaluate the impacts of their ideas. The program is open to all faculty members, postdocs, and graduate students who want to reflect on and improve their teaching.

Key Questions:

- How can your institution incentivize cross-department collaboration across academic departments, disciplines, and faculty to participate in the design, implementation, and evaluation of AI tools that support teaching and learning?
- How can research on social and behavioral sciences enhance the rigorous evaluation of AI systems?

Resources:

- The National AI Research Resource Pilot offers training and computational resources to help educators integrate AI into their curricula, helping them stay current with AI developments. This enables students to acquire practical AI skills for future careers and ensures that institutions can effectively teach and utilize AI technology.
- The EDUCAUSE Higher Education Generative AI Readiness Assessment guides institutions in understanding their preparedness for strategic AI initiatives.
- In developing AI-related infrastructure and platforms, institutions may find it valuable to consult the Department of Labor's Artificial Intelligence and Worker Well-being: Principles for Developers and Employers.
- The Report from the National Institute for Learning Outcomes Assessment, *Equity and Assessment: Moving Towards Culturally Responsive Assessment*, provides guidance on culturally responsive assessments, equitable evaluation methods, and data disaggregation to support targeted interventions for diverse students.
- The University of Maine's Learn with AI Toolkit and Auburn University's self-paced Teaching With AI course provide support for faculty and students to integrate AI into their teaching and learning processes.
- The University of Michigan offers free GenAI services to students, faculty and staff. These include U-M GPT for accessing popular AI models, U-M Maizey for enriching experiences with custom datasets, and U-M GPT Toolkit for advanced AI environment control.

Recommendation 3: Rigorously test and evaluate AI-driven tools, supports, and services. Conduct rigorous research and evaluation studies on AI-driven platforms prior to deployment. Use continuous improvement methods to ensure their effectiveness, safety, and alignment with educational goals and student needs.

Testing and evaluating AI is not monolithic, though its various elements are sometimes conflated. This conflation can cause an institution to overlook important aspects of AI's impact that may have long-term institutional consequences. The Department has included two sub-recommendations in this section to distinguish between two separate – yet related – elements. This distinction helps to ensure testing and evaluation activities meet overall objectives. Recommendation 3a focuses on making sure that AI systems can operate in alignment with ethical and privacy standards, while 3b focuses on using research to build high-quality evidence on the effectiveness (e.g., what works, for whom, and under what conditions) of AI systems in improving education outcomes.

Recommendation 3a. Institutions should test AI products and services to ensure they are safe, equitable, mitigate the risk of algorithmic discrimination, and protect student privacy. These procedures should include rigorous evaluation methods, stakeholder feedback, and ongoing monitoring to ensure AI solutions align with ethical standards, civil rights and privacy obligations, and educational goals.

Implementing robust procedures for testing AI products and services is crucial to mitigate the risk of algorithmic discrimination and ensure that AI-enabled solutions are equitable and protect student privacy. It is recommended that scholars conduct research in partnership with historically underserved groups to promote collaboration among AI technologists and educators to develop culturally responsive AI systems.[10]

In order to understand the potential impacts of AI systems on historically underserved groups, it is important to consider the requirements of federal civil rights laws,[11] including issues of algorithmic discrimination and accessibility as specific equity-related concerns of AI integration. Existing laws (including civil rights laws) are paramount and apply to situations where any variation of artificial intelligence leads to any discrimination across the continuum of a student's learning experience and in all of the school's operations. Federal civil rights laws protect students against discrimination based on race, color, national origin, sex, disability, and age.[12]

---

[10] Mouta, A., Pinto-Llorente, A. M., & Torrecilla-Sánchez, E. M. (2023). Uncovering blind spots in education ethics: Insights from a systematic literature review on artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 33*, 290–324.

[11] Postsecondary institutions and their leaders must comply with civil rights laws and are prohibited from discriminating on prohibited bases against students, applicants, and employees. Institutions may find this overview of the Department of Justice's Civil Rights Division's work on AI and civil rights of interest. Importantly, this guide does not attempt to discuss all of the civil rights provisions that could potentially apply to AI use in education settings.

[12] E.g., Title VI of the Civil Rights Act of 1964, 42 U.S.C. § 2000d, 34 C.F.R. part 100; Title IX of the Education Amendments of 1972, 20 U.S.C. § 1681, et seq., 34 C.F.R. part 106; Section 504 of the Rehabilitation Act of 1973., 29 U.S.C. § 794, 34 C.F.R. part 104; Americans with Disabilities Act of 1990, 42 U.S.C. § 12131, et seq., 28 C.F.R. parts 35 and 36; Age Discrimination Act of 1975, 42 U.S.C. § 6101, et seq., 34 C.F.R. part 110.

The White House Office of Science and Technology's Blueprint for an AI Bill of Rights defines "algorithmic discrimination" and outlines an appropriate institutional response:

> Algorithmic discrimination occurs when automated systems contribute to unjustified different treatment or impacts disfavoring people based on their race, color, ethnicity, sex (including pregnancy, childbirth, and related medical conditions, gender identity, intersex status, and sexual orientation), religion, age, national origin, disability, veteran status, genetic information, or any other classification protected by law. Depending on the specific circumstances, such algorithmic discrimination may violate legal protections. Designers, developers, and deployers of automated systems should take proactive and continuous measures to protect individuals and communities from algorithmic discrimination and to use and design systems in an equitable way.

Institutions should test and evaluate AI products or services used in educational contexts, both proactively and on an ongoing basis, to mitigate the risk of algorithmic discrimination. Research has documented algorithmic discrimination in specific educational applications of AI (or prior uses of machine learning), including when race/ethnicity and socioeconomic status are used to predict a particular outcome (e.g., the likelihood of a student failing a course).[13, 14] The use of AI systems to deliver services or inform decisions could discriminate against students on a prohibited basis, or erode their privacy, and these are issues that can be addressed by developers as early as the initial design phase through the testing and implementation phases. In the spirit of *mitigating the risks associated with AI in order to seize its benefits*, institutions will need to attend to emergent risk such as:

- As guidance counselors use AI-assisted tools to recommend college and career pathways, who will detect and counter unfairness in the recommendations due to biases in historical data sets that were used to develop the AI model, and which could harm vulnerable populations?
- As educators use AI to simplify their work by writing emails or other correspondence about their students' work, who is responsible for safeguards against disclosing a student's private information to unintended recipients, including the developer of the AI model?
- As administrators and school leaders procure early warning systems to identify students who may be "at risk," who has sufficient knowledge and time to evaluate whether the AI developer adhered to scientific, legal, and privacy standards necessary to safeguard students' civil rights?
- As educators deploy anti-plagiarism detectors to identify a student's inappropriate use of edtech, who has responsibility for recognizing weaknesses and biases in AI-based detectors that could lead to disciplining students unfairly or unequally? Who ensures that underserved and vulnerable populations are not unfairly targeted?

---

[13] Gándara, D., Anahideh, H., Ison, M. P., & Picchiarini, L. (2024). Inside the black box: Detecting and mitigating algorithmic bias across racialized groups in college student-success prediction. *AERA Open, 10.*

[14] Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education,* 1-41

To protect against discriminatory outcomes, developers and institutions will need to work together to involve students and faculty from historically underserved groups in every phase of the design and implementation life cycle. This involves designing evaluations of AI tools to be both comprehensive by covering a broad, random sample, and detailed by analyzing the specific impacts on different subgroups to ensure equitable outcomes.

Because educational inequity can occur unintentionally, institutions should focus testing and evaluation on the overall impact of AI. This applies to the use of AI-enabled solutions in curriculum, as well as any technology or solution that can be used for monitoring behavior, classroom management, or discipline.

Note that the potential for algorithmic discrimination is not limited to applications that make big decisions, such as guiding student course or career selection, but can also occur in a series of smaller or seemingly inconsequential decisions (for example, in the pacing or content of technology-based lessons) that, taken together, may create an inequitable learning environment for students. At the same time, it is worth noting that AI systems can also be a valuable asset in the effort to promote equity. For example, AI systems can be designed to challenge and mitigate biases through certain tools that help determine how high-level concepts, such as "race," "gender," or "income," influence the decisions made by AI models. By quantifying the impact of these concepts on a model's output, these tools aid in understanding, detecting, and potentially mitigating biases within AI systems.[15]

Any AI-driven student support tools that are developed or adopted should comply with the Web Content Accessibility Guidelines (WCAG) international standard and any emerging guidelines specific to AI and the web.[16] This ensures accessibility to web content for users, including those with disabilities, by providing equal access to information and functionality. Compliance with these standards also addresses the unique challenges and opportunities of AI technologies. Additionally, these tools should undergo regular evaluations to ensure they continue to meet evolving accessibility standards and best practices.

In addition to equity concerns, the integration of AI in postsecondary education also raises critical issues related to safety, data privacy, and ethical use. Training and improving AI systems can require extensive data collection, storage, and use, including sensitive student information, which necessitate robust privacy safeguards to prevent misuse and ensure compliance with federal laws such as the Family Educational Rights and Privacy Act (FERPA).[17]

One major area of concern is the potential for AI systems to inadvertently infringe on student privacy through the monitoring and analysis of academic data. These systems can track student

---

[15] Christian, B. (2021). The alignment problem: How can machines learn human values? Atlantic Books.

[16] Regulations issued under Title II of the Americans with Disabilities Act, 28 C.F.R. §§ 35.200 – 35.205, establish requirements for the accessibility of State and local public entities' (including public postsecondary educational institutions) web content. For information about the requirements and the effective dates of these requirements, see ADA.gov.

[17] Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education, 20*(22).

performance, engagement, and even behavioral patterns, which, if not properly managed, could lead to unauthorized access or misuse of personal information. To mitigate these risks, institutions should implement stringent data governance frameworks and ensure transparency in how student data is collected, stored, and used.[18]

The Family Educational Rights and Privacy Act (FERPA) (20 U.S.C. § 1232g; 34 CFR part 99) is a federal law that protects the privacy of students' education records. The term "education records" is defined under FERPA, with certain exclusions, as those records that are directly related to a student and which are maintained by an educational agency or institution, such as a college or university, or by a party acting for the agency or institution, to which funds have been made available under any program administered by the Secretary of Education. FERPA affords eligible students certain rights with regard to their education records. Under FERPA, an "eligible student" is any student who is 18 years of age or is attending a postsecondary institution at any age. These rights include the right to inspect and review education records, the right to seek to have the education records amended, the right to have some control over the disclosure of personally identifiable information from the education records, and the right to file a written complaint with the Student Privacy Policy Office in the U.S. Department of Education regarding an alleged violation of FERPA. Under FERPA, an educational institution is prohibited from disclosing personally identifiable information from education records, without consent, unless the disclosure meets an exception to FERPA's general consent requirement. One such exception permits an educational institution to disclose, without consent, personally identifiable information from student education records to contractors, consultants, volunteers, or other third parties to whom the institution has outsourced institutional services or functions, provided that the outside party: constitutes a school official with a legitimate educational interest under the criteria listed in the institution's annual notification of FERPA rights; performs an institutional service or function for which the institution would otherwise use employees; is under the direct control of the institution with respect to the use and maintenance of education records; and is subject to the requirements in FERPA which provides that the personally identifiable information from education records may be used only for the purposes for which the disclosure was made, and which governs the re-disclosure of personally identifiable information from education records. In addition, an educational institution must use reasonable methods, such as physical or technological access controls, to ensure that school officials obtain access to only the education records in which they have legitimate educational interests. For more information on FERPA, visit the Department's Student Privacy website: https://studentprivacy.ed.gov/. The Guidance for Eligible Students is at: https://studentprivacy.ed.gov/resources/eligible-student-guide-family-educational-rights-and-privacy-act-ferpa.

[18] Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., …Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education, 21*(4).

Key Questions:

- What strategies can your institutions use to deploy AI platforms that operate with appropriate levels of ethical human supervision?
- How can your institution incorporate comprehensive stakeholder feedback into the testing and monitoring of AI solutions to ensure these technologies align with ethical standards and educational goals? How will this process ensure that there is robust crosstalk between administrators, developers, and stakeholders most at risk of algorithmic discrimination?
- What testing, ongoing monitoring practices and technical, administrative, or physical safeguards are essential for protecting student privacy and avoiding algorithmic discrimination while using AI-driven products and services? How can institutions ensure continuous compliance with these practices? Similarly, what monitoring practices are needed to identify and assess AI's impact on equity and fairness, and mitigate the risk of algorithmic discrimination when it is present?

Resources:

- The Artificial Intelligence and Algorithmic Fairness initiative aims to ensure that AI and other emerging tools used in employment decisions comply with federal civil rights laws. It will closely examine how technology changes employment decisions and guide applicants, employees, employers, and technology vendors to use these technologies fairly, in line with federal equal employment opportunity laws.
- NIST is working on a standard for identifying bias in AI, with three main ways in which bias occurs (systemic, statistical/computational, and human), which intersect with different foci for mitigating bias (in datasets and models, during development of applications, and as applications of AI are used in the field). NIST has also released a generative AI risk profile that is a companion resource to the AI RMF.
- A Global Information Society Watch 2019 report details examples of opportunities and challenges of AI in different use contexts, including education. The report has a particular focus on social justice and human rights, drawing on examples across the globe in various sectors.

Recommendation 3b. Institutions should evaluate AI systems through iterative cycles of testing, feedback, and improvement. These efforts should build high-quality evidence on the abilities of AI platforms to support the institutions' student services by providing continuous, personalized, and holistic support to address the needs of diverse learners.

Studies of AI-driven tools should determine what works, for whom, and under what conditions. Comprehensive and transparent reporting of these studies, in accordance with applicable law, builds trust and guides evidence-based decision-making in educational policy and practice. This applies to areas such as advising systems, academic supports, mental health resources, and career guidance, to ensure that AI enhances the overall educational experience. In designing and developing these studies, institution leaders may find the standards set forth by the Institute of Education Sciences for evaluating research studies to be a helpful resource.

Deciding on the most appropriate AI model(s) to use for a particular purpose is not always straightforward. Some institutions have developed processing for making these determinations by using a combination of benchmarking studies and rigorous research across different AI models. These studies help institutions identify which models or configuration of models are most effective and under what circumstances. Where possible, and in accordance with applicable law, the results of these studies should be shared with other institutions. For example, the Institute of Education Sciences works to help institution leaders generate, find, and use high-quality evidence through the What Works Clearinghouse and the Regional Educational Laboratories. Furthermore, given the ongoing advancements in AI, it is crucial to continually evaluate whether existing models still perform as expected or if new models or systems need to be developed or integrated.

> Benchmarking studies are ways of measuring the performance of models against a defined set of metrics. These studies are often useful in comparing the technical performance of different models. The result of these studies can help institution leaders to understand broadly the models' capabilities (e.g. error rates on translation, scores on standardized tests, or solving opened physics problems). These studies themselves do not tell you how well the model will perform when deployed in complex or changing environments, and they cannot be used to determine causal impacts of an AI model on student learning. For that, institutions must use rigorous research designs like those described above from the What Works Clearinghouse.

Benchmarking studies and rigorous research of new and existing AI platforms in postsecondary education should be used to evaluate the effectiveness, scalability, and impact of the platforms on student learning outcomes and administrative efficiencies. AI outputs should be monitored through both automated quality assurance systems–for example, a screening tool that reviews user inputs and AI outputs, especially when operating at scale–and experienced human advisors, who can intervene when complex or sensitive issues arise that require human judgment.

Key Questions:

- How can your institutions ensure that human advisors adequately review interactions between students and the AI system to provide feedback and updates to the AI platform? Similarly, how can your institution receive feedback from students on their experience with AI platforms?
- In what ways can your institutions better include diverse voices in AI development and adoption to create an inclusive and equitable AI ecosystem in education?
- How can your institutions create streamlined pathways to facilitate comprehensive evaluation of AI models (e.g., appropriate access to high-quality data or clear reporting procedures of findings)?

Resources:

- Two examples of collaborative infrastructure are the Institute of Education Sciences' SEERNet program and the National Science Foundation's SafeInsights initiative, both of which leverage digital learning system data for education while prioritizing student privacy.

- The Center for Equitable AI and Machine Learning Systems (CEAMLS) at Morgan State University advances research, standards, and technology for fair and unbiased AI. CEAMLS educates students and the public on the impact of AI, serving as an interdisciplinary hub for equitable tech applications.

Recommendation 4: Seek collaborative partners for designing and iteratively testing AI models across educational applications. Forge partnerships with industry, nonprofit organizations, and other postsecondary institutions on AI design and testing, including AI use for enhancing learning experiences, improving institutional processes, and supporting diverse student needs.

Collaborative partnerships between educators, researchers, and technology developers offer benefits to an educational institution (or a consortia of education institutions) in several ways. Institutions benefit from educators' expertise in pedagogy, researchers' expertise in measurement and evaluation, and from technology companies' technical expertise and resources, allowing for the creation of more tailored educational solutions that address specific institutional needs.[19, 20] For example, institutions that work with technology providers may benefit from cost-effectiveness and economies of scale, lessening the financial load of developing technology in-house.[21]

Institutions should structure partnerships so that educators and learners are included throughout the AI development and deployment process and seek to include not only decision makers but also those who will be most affected by design choices in a product or service. Through ongoing collaboration, educational institutions, research entities, and technology partners will be better equipped to enhance learning environments and student support systems and improve operational efficiencies.

Productive partnerships can be fostered through collaborative networks that bring together diverse stakeholders. Here we highlight a few examples of different types of partnerships and their various goals:

- The Institute of Education Sciences' SEERNet program is a network of platform developers, researchers, and education stakeholders who collaborate to conduct research that leverages the large-scale reach of digital learning platforms (DLPs) while also informing the design of learning experiences within DLPs. The collaborative approach facilitates the alignment of DLP improvements with education needs and research findings. Networks like SEERNet enable stakeholders to share best practices, conduct research, and implement advanced technologies to enhance education.
- Arizona State University (ASU) is leveraging partnerships to explore and implement a variety of AI-driven approaches. The ASU AI Innovation Challenge encourages members of the ASU community to propose and develop impactful AI applications across various

[19] Selwyn, N. (2021). *Education and Technology: Key Issues and Debates* (3rd ed). Bloomsbury Publishing.
[20] U. S. Department of Education, Office of Educational Technology. (2017). *Reimagining the Role of Technology in Education: 2017 National Education Technology Plan Update*.
[21] Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher, 41*(1), 16-25.

academic and administrative areas. Faculty across disciplines are exploring the potential of generative AI through ASU's Generative AI Community of Practice.

- The City University of New York (CUNY) and its partners developed a predictive AI tool that increased graduation rates from 54 percent to 86 percent in two years. The AI model used data from students facing various challenges, such as being first-generation or working while studying, and it analyzed 75 risk indicators (e.g., variation in grades, attendance patterns, GPA semester average, years of enrollment) to help advisers support students at risk of dropping out. The project emphasized co-creating solutions, refining models through small-scale trials, and using AI to augment, not replace, personal support.[22]
- The National Center for Student Success at Georgia State University is also working to support a consortium of institutions in leveraging research-based AI chat bots to help improve student success.

Key Questions:

- How can your institution identify and prioritize potential internal and external collaborative partners to effectively design and test AI models that enhance learning experiences and improve institutional processes?
- What best practices should your institution follow to ensure that partnerships in AI design and testing enable work that addresses diverse student needs and promotes equitable outcomes across various educational applications?
- How can your institution effectively advocate for and partner with other institutions and the technology industry to prioritize the development of AI models that are transparent, accessible, and aligned with educational needs?

Resources:

- The Department recently released *Designing for Education with Artificial Intelligence: An Essential Guide for Developers*. This guide is intended to support working organizations developing AI platforms for use in education settings. As a central principle, it encourages and provides recommendations on collaboration between education institutions and developers, emphasizing the importance of designing educational solutions in partnership with educators to build a foundation of trust.
- The Artificial Intelligence Incubator Network Initiative, managed by the American Association of Community Colleges (AACC), aims to establish AI incubators nationwide with industry funding. This 18-month project supports community colleges by developing AI content and providing resources such as monthly discussions, economic development opportunities, student engagement strategies, and AI incubation best practices.
- Lever for Change, an affiliate of the John D. and Catherine T. MacArthur Foundation, provides resources and guidance for both initiating new collaborations and enhancing existing ones. These resources include advice on strategic mergers, key questions for starting

[22] Lucariello, K. (2023, October 12). *AI predictive model partnership dramatically raises CUNY graduation rate.* Campus Technology.

collaborations, and tools for creating inclusive planning processes. The aim is to support effective and equitable partnerships in various sectors.

Recommendation 5: Review, refine, and supplement program offerings in light of the growing impact of AI on future jobs and career opportunities. Institutions may regularly assess current programs and potentially create new programs to equip both students and workers with the skills necessary for a job market increasingly influenced by AI.

The rapid advancement of AI technologies requires a proactive approach in postsecondary education to ensure that academic programs remain relevant and adequately prepare students for the evolving job market. The potential impact of AI on current and future job roles indicates there is a need for a re-appraisal of the relevance and adequacy of postsecondary educational offerings. To support their students, institutions should prepare to reshape instruction to focus on those competencies that are essential for navigating the changing world of work, such as critical thinking, creativity, communication, and collaboration.

Academic leaders should evaluate the competencies required for careers within their disciplines, considering the capabilities of generative AI tools and related advances. This appraisal should also address the potentially positive and negative impacts of AI advances on particular roles and tasks and the arrival of new AI-intensive roles or tasks.

Institutions should also begin equipping students from all disciplines with the AI literacy skills (i.e., the knowledge and skills to understand, use, and evaluate AI systems critically, promoting safety and ethics), and AI specific courses they are likely to need for their careers. This includes placing an appropriate amount of emphasis on the diverse impacts that automation is having on various fields and highlighting those fields that are projected to see both high rates of task-specific automation and an overall increase in demand for workers.[23] For example, the University of Florida's initiative, AI Across the Curriculum, incorporates AI literacy for all students and specialized AI curricula in specific departments.

Another example is Barnard College developing an AI literacy framework that serves as a conceptual guide for AI education in postsecondary education.[24] This framework categorizes AI literacy into four levels: understanding AI, using and applying AI, analyzing and evaluating AI, and creating AI. These categories are reflected in Barnard's curriculum offerings, such as the "GenAI 101" workshops, which provide hands-on training with AI tools, and "Collaborative Prompt Engineering Labs," which focus on refining skills in generating and improving AI outputs. Additionally, Barnard offers instructional workshops on AI ethics, covering critical topics like ethical implications, copyright issues, and environmental impacts.

The importance of integrating AI literacy into educational programs is perhaps most apparent when one examines AI's transformative impact on scientific research. AI accelerates data analysis, facilitating faster processing of vast datasets and significantly reducing the time from hypothesis to

[23] McKinsey Global Institute. (2024, May 21). *A new future of work: The race to deploy AI and raise skills in Europe and beyond.* McKinsey & Company.
[24] Coffey, L. (2024, June 11). *Inside Barnard's pyramid approach to AI literacy*. Inside Higher Ed.

discovery. It also enhances predictive modeling, leading to more accurate simulations in areas such as climate science and medicine. Furthermore, AI fosters interdisciplinary innovation by automating routine tasks, allowing scientists to focus on creative problem-solving, thus driving the next wave of scientific breakthroughs.

One compelling example of AI's impact is in pharmaceutical research, where AI is transforming the drug discovery process.[25] Traditionally, drug discovery has been costly and time-consuming, with high failure rates. However, AI-driven platforms can now identify novel drug targets for complex diseases and design entirely new molecules that would have been extremely difficult for human chemists to identify. AI can test and optimize hundreds of potential drug candidates in a fraction of the time required by traditional methods, increasing the efficiency and success rates of drug development. Once a potential drug target is identified, AI generates a large array of candidate molecules and tests them in simulated environments to assess their efficacy and safety. This iterative process allows for the refinement of molecules, informed by data from previous studies, patient records, and molecular simulations. The result is a streamlined drug discovery process that explores multiple hypotheses simultaneously, often leading to the discovery of novel and structurally unique molecules, thus expanding the boundaries of chemical and biological research.

The integration of AI across contexts exemplifies the evolving landscape that colleges and universities will need to monitor, so they can prepare their graduates to enter a workforce with adaptive skills and proficiency in leveraging tools for automation. As AI continues to evolve, its role in education and research will be critical in preparing the next generation of professionals to navigate and shape the future of their respective fields.

Key Questions:

- How can your institutions keep abreast of emerging demands for new AI-enabled professions?
- What steps can your institutions take to make sure students from historically underserved groups are aware of, and have opportunities to enter, new programs centered around AI-enabled competencies?
- How can AI be integrated into all disciplines to enhance the relevance of these programs in an AI-driven job market?
- How can your institutions use AI to scale effective lifelong learning and training models, and what role can AI play in advancing apprenticeships and skill development for displaced workers?
- What metrics and evaluation methods can be used to assess the effectiveness of AI-integrated programs and their alignment with job market requirements?

Resources:

- Ohio State offers an online Artificial Intelligence in Digital Health certificate that explores AI's role in the health care industry.

[25] Arnold, C. (2023). Inside the nascent industry of AI-designed drugs. *Nature Medicine, 29*(8), 1292–1295.

- Emory University unveils interdisciplinary AI minor open to all undergraduates.
- Purdue University's College of Liberal Arts has introduced an interdisciplinary program that blends philosophy with computer science, equipping students with the programming and data analysis skills necessary to navigate the ethical challenges posed by AI.
- Miami Dade College opened an AI Center that provides resources to students, faculty, staff, and the community across all campuses to teach, learn, and collaborate on AI topics.
- The MIT Center for Information Systems Research collaborated with Johnson & Johnson to develop MySkills, a machine learning-based talent development platform. MySkills assesses employees' current skill levels, identifies skill gaps, and offers development and career opportunities.
- The Vets in AI program integrates veterans into the AI field by providing education, employment, and entrepreneurship opportunities. It offers an extensive AI curriculum, bootcamps, workshops, and mentorship, and fosters collaboration between veterans, tech companies, and policymakers.
- Jobs for the Future published *The AI Ready Workforce*, which evaluates the potential impact of AI on the tasks and skills associated with jobs across various sectors of the economy.
- The Council of Economic Advisors in The White House offers an economic framework for understanding the impacts of AI on the workforce in Chapter 7 of the Economic Report to the President.
- The National Academies published *Artificial Intelligence and the Future of Work*, which explores the rapid advancements in AI and their profound impact on the workforce, productivity, and various industries. It addresses AI's opportunities in enhancing work and innovation, alongside challenges like bias and societal integration.

Conclusion

The integration of AI in postsecondary education offers significant opportunities to enhance learning outcomes, streamline institutional operations, and prepare students for a future that is increasingly defined by fast-paced technological change. This brief has outlined the transformative potential of AI in various aspects of colleges and universities, including admissions, enrollment, academic advising, and personalized learning environments. By leveraging AI, institutions can improve efficiency, support data-driven decision-making, and create more personalized and engaging student experiences.

However, the deployment of AI in postsecondary education should be approached with careful consideration of ethical implications, particularly concerning equity, fairness, privacy, and the non-discriminatory use of technology. Institutions have a dual responsibility: to utilize AI to drive institutional improvements that promote broad and equitable access to education and to equip students with the skills necessary to thrive in an AI-influenced world. This entails developing robust data governance policies, investing in AI literacy and professional development, conducting research on AI's impact on student learning, and ensuring that AI applications are safe, transparent, accountable, and aligned with educational goals.

The recommendations provided in this brief offer a strategic framework for postsecondary education leaders to navigate the complexities of AI integration. Key areas of focus include establishing transparent policies, building infrastructure to support AI innovation, conducting rigorous research on AI tools, fostering collaborative partnerships, ensuring equitable and ethical AI use, and regularly updating academic programs to align with the evolving job market.

As AI continues to evolve, it is imperative for postsecondary institutions to remain adaptable and proactive in their approach. By doing so, they can harness the potential of AI to enhance educational quality, support student success, and contribute to a more inclusive and equitable society. The Department remains committed to supporting institutions in this journey, providing guidance and resources for the safe, responsible, and non-discriminatory use of AI in postsecondary institutions.

Addendum on Evidence-Based Insights About AI Integration

This addendum has two major subsections: AI in Enhancing Learning and Instruction, and AI Support for Institutional Operations. This section conveys evidence-based insights about the integration of AI in institutional academics and operations to support institutional leaders' decision-making on policies and supports. This non-exhaustive literature review is intended to assist researchers and institution leaders to better understand the research base associated with the use of AI systems by postsecondary education institutions. The studies summarized in this addendum provided the foundation for crafting the brief's recommendations; however, it is also important to note that the Department considered input from a variety of sources in crafting the recommendations.

AI in Enhancing Learning and Instruction

AI technologies are likely to profoundly impact teaching, learning, and assessment. Regardless of whether an institution has a formal strategy for incorporating AI into academics, generative AI tools are already widely used.[26] By understanding both AI's capabilities and its shortcomings, academic leaders and administrators can position their institutions to use AI in teaching and learning in ways that complement human interactions and support students' academic success. In this brief, we have focused our attention on topics and studies that address several priority aspects of postsecondary education. For a more comprehensive review of AI and its impact on instruction broadly, readers may find it valuable to consult the Department's *Artificial Intelligence and Future of Teaching and Learning Report*.

AI-Enhanced Learning Environments

A substantial research base supports the use of AI-driven adaptive learning environments for enhancing learning outcomes. Adaptive systems are built on learner models that consider what the learner knows and does, addressing cognitive, affective, and behavioral aspects of learning.[27] These systems consider variables such as learner knowledge and proficiency (cognitive), interest in the topic (affective), and study habits and help-seeking behaviors (behavioral). A review of studies of the efficacy of AI-enabled adaptive systems found significant improvements in students' cognitive learning outcomes compared to non-adaptive interventions, particularly in computer science and mathematics courses. The effect sizes varied, with longer interventions showing greater impacts.[28]

[26] Tyton Partners. (2023, October). GenAI in Higher Education: Fall 2023 Update Time for Class Study.

[27] Vandewaetere M., Desmet P., & Clarebout G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. Computers in Human Behavior, 27(1), 118–130.

[28] Wang, X., Huang, R., Sommer, M., Pei, B., Shidfar, P., Rehman, M. S., …Martin, F. (2024). The efficacy of artificial intelligence-enabled adaptive learning systems from 2010 to 2022 on learner outcomes: A meta-analysis. *Journal of Educational Computing Research, 0*(0).

The personalization in AI-driven adaptive learning environments is key to their success. AI-based systems provide content to students at appropriate levels of difficulty based on the learners' past results. This is particularly salient for providing students with feedback on their learning. Research shows that low-stakes quizzes and frequent practice tests with prompt feedback improve learning outcomes by reinforcing material and identifying knowledge gaps. Timely feedback helps correct misunderstandings quickly, enhancing retention and comprehension.[29, 30] Formative feedback is also enhanced, as these systems can track student progress over time, identifying trends and patterns in performance to offer continuous feedback that evolves with the learner's development.[31]

Generative pretrained transformer (GPT) models are proving capable of providing just-in-time individualized help for cognitive aspects of learning. A randomized study compared learning gains from a GPT to those from human tutor-authored help across four math subjects. Results showed that learners receiving GPT help had significantly higher gains than learners in the control condition who received no help at all. Learners receiving human tutor-authored help also showed gains, but the gains were not statistically distinct from those for the control or GPT groups. GPT-generated solutions were produced 20 times faster than human-authored solutions, including time for error checking the GPT output. High "hallucination"[32] or "confabulation" rates in early versions of the GPT help were reduced to nearly 0 percent in Algebra and 13 percent in Statistics by using self-consistency, an error mitigation technique. Near-zero hallucination rates suggest the potential for dynamically generating personalized help in Algebra without manual error checking.[33]

Educators can also benefit from integrating AI into learning environments. AI can automate routine tasks, allowing instructors to focus their time and effort on areas where they can add the most value. For instance, in adaptive learning environments AI systems can provide students with immediate feedback, which enables educators to offer support where it is most needed.

AI-enabled tools such as essay scoring systems and Automatic Short Answer Grading (ASAG) are becoming more common. These systems use AI to assist in examining the content of essays to score the quality of students' writing, comprehension of topics, and ability to express ideas clearly. ASAGs are designed to calculate how closely a student's response aligns with the expected or model answer to grade concise written replies typically found in short-answer exams. A review of ASAGs describes the variability in the accuracy of these systems, with some achieving strong correlations with human scores and others indicating significant room for improvement. The review highlights

---

[29] Haughney, K., Wakeman, S., & Hart, L. (2020). Quality of feedback in higher education: A review of literature. Education Sciences, 10(3), 60.

[30] Quality Matters. (2023). *Specific Review Standards from the Quality Matters Higher Education Rubric*. (7th ed.)

[31] Luckin, R. (2017). Towards artificial intelligence-based assessment systems. *Nature Human Behaviour, 1*(0028).

[32] This term has been used in popular discussion about generative AI. The official term that is adopted by the National Institute for Standards and Technology is "confabulation."

[33] Pardos, Z. A., & Bhandari, S. (2024). ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills. *PLoS ONE, 19*(5), e0304013.

that ensemble methods have shown significant promise in enhancing the robustness and effectiveness of ASAG systems.[34]

> The ensemble approach enhances prediction accuracy by combining multiple models. For example, in predicting student dropouts in online courses, it uses different prediction methods to aggregate results for greater precision. A random forest classifier, an ensemble method, consists of numerous decision trees analyzing different student data aspects, such as grades, participation, and login frequency. Each tree makes its own prediction, with some predicting retention and others predicting dropout. The random forest then takes the majority vote among the trees, providing a more reliable and accurate prediction than any single tree could have. This method helps online platforms better identify students needing extra support to stay in the course.

Findings from the National Assessment of Educational Progress (NAEP) Math and Reading Automated Scoring Challenges indicate promising outcomes for automated scoring systems. The goal of the NAEP challenge was to assess the potential of automated scoring techniques in evaluating open-ended responses to reading assessment items, with a focus on determining the current capabilities, accuracy metrics, validity evidence, and cost efficiencies of an automated approach. Key findings highlight that automated scoring can effectively replicate human scores assigned to most NAEP reading items, with large language models proving to be the most accurate approach in scoring assessments. It was found that the results in the math challenge did not exhibit bias, in contrast to the reading challenge where some items showed significant bias for the English learner (EL) subpopulation.

AI can automate time-consuming aspects of teaching and provide feedback to instructors on their instructional practices, helping them improve. A recent study examined the effectiveness of an AI tool designed to support instructors in a large-scale, online programming course, which facilitated weekly meetings between instructors and students conducted on an online video platform. The tool provides instructors with weekly automated feedback on incorporating student contributions, such as acknowledging, revoicing, and integrating student ideas into instruction. The results showed that instructors who received the feedback showed improved integration of student ideas, enhancing instructional quality and student engagement. However, the study authors noted decreased accuracy in the speech transcriptions for non-native English speakers and that the effectiveness of the feedback varied with instructors' characteristics, such as their gender and teaching experience in the course.[35] These issues could raise potential concerns under federal civil rights laws as described under Recommendation 3a.

---

[34] Gao, R., Merzdorf, H. E., Anwar, S., & Hipwell, M. C. (2024). Automatic assessment of text-based responses in post-secondary education: A systematic review. *Computers and Education: Artificial Intelligence, 6*, 100206.

[35] Demszky, D., Liu, J., Hill, H. C., Jurafsky, D., & Piech, C. (2023). Can automated feedback improve teachers' uptake of student ideas? Evidence from a randomized controlled trial in a large-scale online course. *Educational Evaluation and Policy Analysis, 0(0).*

Note that AI technologies can also be embedded in or support parallel technologies. For example, the synergies between AI and extended reality (XR) technologies, including virtual reality (VR) and augmented reality (AR), are likely to grow in the near future. Recent studies on the use of XR technologies in educational contexts reveal significant enhancements in skills development and learner engagement, particularly in hospitality, medicine, and science studies.[36] For instance, a study involving engineering students demonstrated that those who used an AR intervention had significantly higher posttest operational skills (e.g., connecting equipment, generating signals, adjusting controls) compared to the control group, underscoring the positive impact of AR on skills acquisition.[37]

There is also growing interest in designing XR platforms to support students with disabilities. A meta-analysis evaluating the effectiveness of VR- and AR-based interventions for individuals diagnosed with autism spectrum disorder found significant positive outcomes. Studies on VR training and rehabilitation found large effect sizes in improving individuals' daily living skills, with moderate effect sizes noted in enhancing cognitive skills, social-emotional regulation skills, and communication skills. Studies on AR-based training also showed promising effects, although these applications require further research.[38]

Several studies of an immersive VR classroom simulation program augmented with generative AI have found that the simulation effectively bridges the gap between theory and practice for pre-service teachers.[39, 40] Other scholars have offered a framework for developing educational content using XR technologies for adult learners, focusing on complex skills training across various professional fields. The framework combines learning theories with learners' proficiency levels to define competency-based learning objectives and activities. It incorporates XR form factors (various types of XR technologies suitable for learning and training), and learning loops, which are structured sequences of steps to facilitate effective learning processes. The framework enhances the effectiveness of XR training programs by providing a roadmap for designing XR applications that improve adult learning, making the training process more efficient and effective.[41]

---

[36] Bermejo, B., et al. (2023). AR/VR teaching-learning experiences in higher education institutions (HEI): A systematic literature review. *Informatics, 10*(2), 45

[37] Singh, G., & Ahmad, F. (2024). An interactive augmented reality framework to enhance the user experience and operational skills in electronics laboratories. *Smart Learning Environments, 11*(5).

[38] Karami, B., Koushki, R., Arabgol, F., Rahmani, M., & Vahabie, A.-H. (2021). Effectiveness of virtual/augmented reality–based therapeutic interventions on individuals with autism spectrum disorder: A comprehensive meta-analysis. *Frontiers in Psychiatry, 12,* 665326.

[39] Dieker, L., Hughes, C., & Hynes, M. (2023). The past, the present, and the future of the evolution of mixed reality in teacher education. *Education Sciences, 13*(11), 1070.

[40] Lee at al. (2023). *Generative agent for teacher training: Designing educational problem-solving simulations with large language model-based agents for pre-service teachers*. NeurIPS'23 Workshop on Generative AI for Education (GAIED).

[41] Stanney, K. M., Skinner, A., & Hughes, C. (2023). Exercisable learning-theory and evidence-based andragogy for training effectiveness using XR (ELEVATE-XR): Elevating the ROI of immersive technologies. *International Journal of Human–Computer Interaction, 39*(11), 2177–2198.

NIST provides guidance on addressing and managing risks associated with bias in the design, development, and use of AI technologies. The guidance assists organizations in designing, developing, deploying, or evaluating AI in various contexts. This comprehensive approach is intended to mitigate the inherent biases in technology processes that can lead to harmful impacts, fostering trustworthiness and public trust in AI.

Incorporate AI Education Across Disciplines

AI will have significant impacts on the workplace in nearly every field by automating tasks, increasing productivity, and reshaping job roles. According to the World Economic Forum's *Future of Jobs 2023* report, 34 percent of job tasks are automated by AI today, and this figure is expected to rise to 42 percent by 2027. A Cornell University study evaluating the potential impact of Large Language Model-enhanced software on the U.S. workforce estimates that approximately 80 percent of the U.S. workforce could have at least 10 percent of their job tasks influenced by LLM-enhanced software, with about 19 percent of workers potentially experiencing AI's impact on at least 50 percent of their job tasks. The researchers conclude that AI impacts are not confined to high-productivity industries, and access to LLMs could enable about 15 percent of all tasks performed by U.S. workers to be completed significantly faster while maintaining quality.[42] These impacts extend well beyond the careers that have historically been impacted by automation and include fields such as the arts and humanities, the law, STEM careers, and community services.

Postsecondary institutions have a crucial role in preparing students for a dynamic workforce and aiding in the reskilling of workers displaced by AI advancements. Previous research confirms the importance of aligning college majors with labor market opportunities to enhance earnings and reduce unemployment rates.[43, 44] AI, as a topic of study, is becoming priority content, not simply for computer scientists and engineers but for all students. Enhancing the alignment between education environments and the evolving workforce landscape can be particularly beneficial for Black, Hispanic, and foreign-born workers, potentially helping them bridge earnings and employment gaps. Researchers stress the role of educational and career counseling services to guide individuals toward high-demand majors, and advocate for programs that encourage students from underrepresented groups to pursue these fields. Employers benefit, in turn, by hiring individuals whose education aligns with job requirements, promoting diversity and inclusivity.[45]

---

[42] Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). *GPTs are GPTs: An early look at the labor market impact potential of large language models*. arXiv.

[43] Chau, H., Bana, S. H., Bouvier, B., & Frank, M. R. (2023). Connecting higher education to workplace activities and earnings. PLoS ONE, 18(3), e0282323.

[44] Sublett, C., & Tovar, J. (2021). Community college career and technical education and labor market projections: A national study of alignment. *Community College Review, 49*(2), 177-201.

[45] Holzman, B., Han, J., Cortes, K. E., Lewis, B., & Chukhray, I. (2024). *Closing the gap for racial minorities and immigrants through school-to-work linkages and occupational match*. (EdWorkingPaper: 24-947). Annenberg Institute at Brown University.

AI Supports for Institutional Operations

In addition to its impacts on what students need to learn and on how teaching and learning unfold, AI technologies are unleashing a host of opportunities to improve the way colleges and other postsecondary educational institutions operate. Areas of active innovation and experimentation with AI include recruiting, admissions, enrollment, academic advising, and support services of all kinds.

AI-Supported Recruiting, Admissions, Retention, and Enrollment Services

AI is supporting recruiting, admissions, retention, and enrollment services in postsecondary education, offering algorithmic tools to streamline processes and enhance decision-making. By incorporating AI technologies, institutions can improve efficiency, personalize student experiences, and make data-informed decisions to support diverse and inclusive student bodies.

Among an institution's strategic operations, forecasting enrollment and attrition is pivotal because it deeply influences institutional planning, resource allocation, and student support services. Although most colleges and universities have been using algorithms to predict enrollment and attrition for some time, recent developments in machine learning are paving the way for institutions to employ computational methods that are more closely aligned to their strategic priorities. By continually refining their prediction models, colleges and universities can adapt to changing educational landscapes and help students navigate these critical transitions.

Researchers have proposed various machine learning models aimed at predicting student attrition in university environments. These models have consistently identified students at risk of dropping out with higher levels of accuracy than other statistical methods.[46, 47] These studies also indicate that incorporating both quantitative and qualitative variables enhances the predictive accuracy of a model. For example, data on study habits as well as GPA and other factors enable these models to discern patterns that might not be apparent through traditional analyses.[48] Other research found that machine learning methods outperformed traditional statistics and even a commercial service in predicting student enrollment, providing highly predictive results with minimal use of demographic data. The careful minimization of the use of demographic is a very important step in mitigating the risk of algorithmic discrimination. Results from multiple years of enrollment predictions proved to be highly accurate and consistently reliable.[49]

---

[46] Barramuño, M., Meza-Narváez, C., & Gaálvez-García, G. (2022). Prediction of student attrition risk using machine learning. *Journal of Applied Research in Higher Education, 14*(3), 974–986.

[47] Kemper, L., Vorhoff, G., & Wigger, B. U. (2020). Predicting student dropout: A machine learning approach. *European Journal of Higher Education, 10*(1), 28–47.

[48] Ahmad, Z., & Shahzadi, E. (2018). Prediction of students' academic performance using artificial neural network. *Bulletin of Education and Research, 40*(3), 157–164.

[49] Hansen, D. M. (2020). Using artificial neural networks to predict matriculation of university prospects. *Strategic Enrollment Management Quarterly, 8*(1), 21–22.

AI is being used to help distribute financial aid in a more balanced way. At one institution, researchers developed an AI-based program to optimize scholarship and financial aid distribution, balancing enrollment and revenue goals with student needs for affordability and accessibility. The algorithm iteratively adjusted the initial fund distribution based on feedback, refining the approach to meet multiple objectives. Using data from domestic first-time freshmen, including test scores, GPA, and financial need, while excluding demographic information to reduce bias, the researchers were able to identify several budget-conscious strategies. These strategies maintained educational affordability, increased access, and can be adapted to meet a variety of specific goals and student needs.[50]

> Using predictive models for need-based aid carries significant risks. Financial aid is crucial for accessing postsecondary education and reducing inequality, enabling individuals from diverse financial backgrounds to pursue their educational goals.[51] Colleges using AI for financial aid allocation should proceed with caution, thoroughly documenting all data processes, examining how aid allocations affect students' graduation rates, and remaining vigilant to potential inequities in student outcomes.

AI-based systems have started to impact the freshman admissions process. Admission offices have traditionally relied on standardized test scores to categorize large applicant pools and identify subsets of candidates for further review. However, this method is subject to biases in test scores and the selection process involved in test-taking. One study investigated a machine learning-based approach that replaces the use of standardized tests in creating applicant subsets by instead considering a broader range of factors derived from student applications, thus enabling a more holistic review. An evaluation of that approach found that the prediction model, trained on previous admissions data, outperformed an SAT-based heuristic and closely mirrors the demographic composition of the last admitted class. The study authors emphasize the crucial role of human advisors in the admission process, highlighting their necessity for providing a nondiscriminatory evaluation. Predictions can help uncover a broader talent base by allowing admissions officers to target a subset of applicants for closer examination, highlighting candidates who might have otherwise been overlooked.[52]

As AI becomes increasingly integrated into learning environments, it is essential to align it with equity and inclusion values to mitigate biases and meet non-discrimination requirements. Although much has been written about AI fairness, practical guidance for practitioners remains scarce. A

---

[50] Phan, V., Wright, L., & Decent, B. (2022). Optimizing financial aid allocation to improve access and affordability to higher education. *Journal of Educational Data Mining, 14*(3).

[51] Dynarski, S., Page, L. C., & Scott-Clayton, J. (2022, July). College Costs, Financial Aid, and Student Decisions (Working Paper No. 30275). National Bureau of Economic Research.

[52] Lee, H., Kizilcec, R. F., & Joachims, T. (2023). Evaluating a learned admission-prediction model as a replacement for standardized tests in college admissions. In Proceedings of the 2023 ACM Conference on Learning at Scale (L@S '23) (pp. 1-10). Association for Computing Machinery.

recent study[53] directly addresses this gap by developing a step-by-step framework for implementing AI fairness techniques, focusing on a grade prediction case study at a public university. The study emphasizes the importance of incorporating multiple and overlapping bases for identification—such as the overlap of race and gender—not merely as data points but as central to institutional values of equity and inclusion. The research shows that techniques like adversarial learning[54] can effectively reduce biases, particularly in overlapping categories like race-gender and race-income. This framework offers actionable insights for practitioners to design more equitable AI systems and helps ensure compliance with relevant policies and regulations.

AI-Supported Student Advising

AI should support the emerging consensus regarding academic advising, as reflected in the Department's Attaining College Excellence and Equity Advising Summit, which emphasized the importance of cohesive advising models based on four key principles: data-driven advising tailored to individual needs, holistic support integrating various forms of guidance, professional development for advisors, and systemic integration linking advising with institutional support services and academic departments. This strategy addresses students' diverse needs and embeds advising within the broader educational framework, creating an environment conducive to student learning and success.

Postsecondary institutions are considering various ways AI can enhance student advising. These include offering personalized academic guidance, integrating career planning with academic advising, and making predictions of students' course performance. Recent studies[55] have examined AI's role in supporting robust college advising systems. AI agents, such as chatbots and recommender systems, can personalize student support, but their effectiveness varies.

> AI technologies, including chatbots, GPTs, and machine learning models, show promise for improving student success. Chatbots are particularly effective in assisting students with immediate administrative tasks, particularly in the early weeks of a semester.[55] GPTs provide better guidance on students' major selection when more details about the majors are included.[56] Additionally, machine learning models that include standardized test scores with administrative data have greatly improved predictive accuracy of identifying students at risk of course failure.[57]

---

[53] Mangal, M., & Pardos, Z. A. (2024). Implementing equitable and intersectionality-aware ML in education: A practical guide. *British Journal of Educational Technology, 00*, 1–36.

[54] Adversarial learning trains machine learning models using deceptive inputs, called adversarial examples, to improve their reliability by exposing them to challenging inputs during training.

[55] Page, L. C., Meyer, K., Lee, J., & Gehlbach, H. (2023). *Conditions under which college students can be responsive to nudging.* (EdWorkingPaper No. 20-242). Annenberg Institute at Brown University.

[56] Lekan, K., & Pardos, Z. (2023). AI-augmented advising: A comparative study of ChatGPT-4 and advisor-based major recommendations. *Proceedings of Machine Learning Research, 38*, 1-20.

[57] Bertolini, R., Finch, S. J., & Nehm, R. H. (2021). Testing the impact of novel assessment sources and machine learning methods on predictive outcome modeling in undergraduate biology. *Journal of Science Education and Technology, 30*(1), 193–209.

AI-Enhanced Student Support

Postsecondary education is starting to explore the many ways in which AI can support students. These include supports for learning how to learn, implementing universal design principles that accommodate student learning differences, and mental health support.

Several studies on AI in postsecondary education highlight the potential of AI in enhancing students' self-regulated learning (SRL) and improving educational outcomes. One study of an AI framework for SRL demonstrated that AI can enhance students' self-evaluation, self-regulation behaviors, and self-efficacy, leading to higher learning gains and satisfaction. This framework helps monitor student behavior, provide feedback, and support the development of SRL skills.[58] A study exploring students' perceptions of AI applications found that students generally perceive AI as beneficial for metacognitive, cognitive, and behavioral regulation but prefer human support for motivational aspects of learning.[59] Other research found that AI-driven dashboards can improve students' academic performance and self-regulation skills and explanations for performance declines, yet a noted limitation was the lack of time management functionality.[60, 61]

AI tools can provide much-needed support for English language learners during their postsecondary education. A generative AI-based video chatbot can act as a speaking partner for language learners through immersive video calls with AI avatars.[62] Intelligent personal assistants using machine learning have been found to positively impact learners' speaking skills and learning attitudes, especially when used with instructor guidance.[63] For writing assignments, AI-based web applications can offer more structured assistance than traditional word processors by reducing learners' cognitive barriers (e.g., working memory) in writing. These tools assist with low-level tasks such as word production and translation, allowing learners to focus on higher-level writing tasks like organization and revision.[64] Automated feedback systems (AFSs) are also valuable, helping students identify areas for improvement, track progress, and gain confidence, which fosters

[58] Huang, X., Dong, L., Vignesh, C., & Kumar, P. (2022). Self-regulated learning and scientific research using artificial intelligence for higher education systems. *International Journal of Technology and Human Interaction, 18*(7), 15.

[59] Shin, I., Im, K., Yoo, M., Roll, I., Seo, K. (2023). Supporting students' self-regulated learning in online learning using artificial intelligence applications. *International Journal of Educational Technology in Higher Education, 20,* 37.

[60] Afzaal, M., Zia, A., Nouri, J., & Fors, U. (2024). Informative feedback and explainable AI-based recommendations to support students' self-regulation. *Technology, Knowledge and Learning, 29*(1), 331–354.

[61] Prasad, P., & Sane, A. (2024). A self-regulated learning framework using generative AI and its application in CS educational intervention design. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2024),* (pp. 1070-1075). ACM.

[62] Wan, Y., & Moorhouse, B. L. (2024). Using Call Annie as a generative artificial intelligence speaking partner for language learners. *RELC Journal, 0(*0).

[63] Yang, C. T. Y., Lai, S. L., & Chen, H. H. J. (2022). The impact of intelligent personal assistants on learners' autonomous learning of second language listening and speaking. *Interactive Learning Environments,* 1–21.

[64] Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing assistant's impact on English language learners. *Computers and Education: Artificial Intelligence, 3*, 100055.

learner autonomy and emotional self-regulation. AFSs are particularly effective when combined with human feedback and appropriate course structure.[65]

Research shows that AI-driven tools can support college students with disabilities, including sensorimotor disabilities. For example, speech-to-text and text-to-speech tools can assist students who have a hearing impairment or who struggle with reading comprehension.[66, 67] Another study used multimodal data and machine learning to detect and influence students' affective states to enhance engagement and learning for students with intellectual impairments. The study found that tailoring activities to individual emotional states significantly increased engagement and reduced boredom, but longer exposure is needed to determine the impact on learning outcomes.[68] A study of neurodivergent learners in higher education noted the lack of alignment between traditional learning models and the unique needs of neurodivergent students. The study highlighted the importance of offering a more flexible and self-directed learning environment and aligning teaching approaches with neurodiversity-affirming practices.[69] Applications of AI hold promise for exploring tailored learning models for supporting neurodivergent learners.

AI-driven tools also have shown promise for supporting students' mental health. A meta-analytic study of AI-driven conversational agents found that use of the tool significantly reduced symptoms of depression and distress among college students. These effects were more pronounced in agents that were multimodal (voice + text), generative AI-based, integrated with instant messaging apps, and targeting clinical, subclinical, and older populations. Students' experience of the tool was shaped by the therapeutic relationship, the quality of the content, and how often they experienced communication breakdowns. The benefits of using AI-based conversational agents are not without risks. Their use should be accompanied by professional oversight to ensure appropriate interventions and mitigate risks such as privacy violations, biases, and safety concerns.[70]

Institution leaders should carefully monitor AI platforms that are intended to support students to prevent and address biases that could lead to unfair or discriminatory outcomes. A systematic literature review examined the ethical dimensions of AI in education, highlighting gaps in studies and emphasizing the need for inclusive AI systems that address the biases that emerge from a Western and STEM-focused approach. The researchers recommended collaboration between AI

---

[65] Li, L., & Kim, M. (2024). It is like a friend to me: Critical usage of automated feedback systems by self-regulating English learners in higher education. *Australasian Journal of Educational Technology, 40*(1), 1–18.

[66] Bakken, J. P., Uskov, V. L., & Varidireddy, N. (2019). Text-to-voice and voice-to-text software systems and students with disabilities: A research synthesis. In V. Uskov, et al. (Eds), S*mart Education and e-Learning 2019. stemSmart Innovation, Systems and Technologies, 144.* Springer, Singapore.

[67] Wood, S. G., Moxley, J. H., Tighe, E. L., & Wagner, R. K. (2018). Does use of text-to-speech and related read-aloud tools improve reading comprehension for students with reading disabilities? A meta-analysis. *Journal of Learning Disabilities, 51*(1), 73–84.

[68] Standen, P. J., Brown, D. J., Taheri, M., Trigo, M. J. G., Boulton, H., Burton A., …Hortal, E. (2020). An evaluation of an adaptive learning system based on multimodal affect recognition for learners with intellectual disabilities. *British Journal of Educational Technology, 51*(5), 1748–1765.

[69] Friedman, Z. L., & Nash-Luckenbach, D. (2024). Has the time come for heutagogy? Supporting neurodivergent learners in higher education. *Higher Education*, *87*(1), 1905–1920.

[70] Li, H., Zhang, R., Lee, Y., Kraut, R. E., & Mohr, D. C. (2023). Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *npj Digital Medicine, 6*(236).

researchers and educators to develop culturally responsive AI systems. The review stressed incorporating ethical AI education into curricula to enhance students' understanding of AI's functions, impacts, and ethics by integrating AI concepts across subjects and promoting societal discussions. It also emphasized maintaining student diversity and agency, suggesting that effective personalization should balance tailored content with opportunities for exploring diverse perspectives and employing independent learning methods and project-based strategies.[71]

Some institutions, especially Tribally Controlled Colleges and Universities, have begun to explore options for leveraging AI-based tools (e.g., language learning applications and digital archives) to support recording, transmitting, and revitalizing Indigenous knowledge rooted in centuries-old traditions. The unique characteristics of Indigenous data, shaped by distinct cultural, social, and historical backgrounds, require careful consideration in ownership, collection, interpretation, and use. AI systems should be designed with cultural sensitivity to respect Indigenous worldviews and avoid biases. Ethical challenges, including potential misuse and misappropriation, demand cautious and respectful approaches. Ensuring data sovereignty and protecting the privacy and intellectual property rights of Tribal communities are critical, as they advocate for the right to own, control, and govern their data. Informed consent and robust data security measures are essential to safeguard Indigenous cultural information.[72]

Increased concerns about campus safety have prompted some institutions to implement advanced AI surveillance technologies, such as facial recognition and geolocation tracking, to enhance security measures.[73] However, collecting these highly sensitive forms of data can result in significant harm if, for example, it results in cases where individuals are egregiously or systematic misidentified, leading to unnecessary interaction with law enforcement or other legal, social, or financial consequences.[74, 75] The collection and storage of Personally Identifiable Information (PII) present substantial privacy concerns, as cameras record the images and activities of individuals. Organizations, including security firms and transportation agencies, may need access to surveillance data, which can lead to misuse or illegal use of shared video records. Privacy-preserving algorithms and systems should be prioritized to avoid ethical challenges and biases.[76]

---

[71] Mouta, A., Pinto-Llorente, A. M., & Torrecilla-Sánchez, E. M. (2023). Uncovering blind spots in education ethics: Insights from a systematic literature review on artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 33,* 290–324.

[72] United Nations Educational, Scientific and Cultural Organization. (2023). *Open data for AI: what now?*

[73] Sasirekha, V., Malarvizhi, C., Ramadevi, R., Mohankumar, N., & Aarthi, P. (2024, March 1-3). Intelligent campus safety management using IoT and CNNs for surveillance, access, and emergency response. Paper presented at the *3rd International Conference for Innovation in Technology (INOCON)*, Karnataka, India.

[74] Weinstein, M. (2020). School surveillance: The students' rights implications of artificial intelligence as K-12 school security. *North Carolina Law Review, 98*(2), 438.

[75] Hassanin, N. (2023, August 23). *Law professor explores racial bias implications in facial recognition technology*. University of Calgary News.

[76] Ardabili, B. R., Pazho, A. D., & Noghre, G. A. (2023). Understanding policy and technical aspects of AI-enabled smart video surveillance to address public safety. *Computational Urban Science, 3,* 21.

Infrastructure to Support AI Innovation

Infrastructure in postsecondary education plays a crucial role in supporting technological innovation.[77] Strategic long-term planning and comprehensive infrastructure are essential to support sustainable AI scaling, adapt to technological advancements, and maintain data integrity and privacy.

According to the EDUCAUSE 2024 AI Landscape Study, training is the most common component of an institution's AI-related strategy, with 56 percent of institutions providing AI training for faculty, 49 percent for staff, and 39 percent for students. However, the proportion of institutions with plans for technological infrastructure is much lower. Only 15 percent of survey respondents reported creating technology infrastructure to run generative AI models on-campus or through infrastructure-as-a-service cloud computing platforms. The cautious approach toward investing in in-house AI capabilities is primarily motivated by concerns over costs, data privacy, and security.

Resilient infrastructure requires establishing institutional policies governing the transparent and ethical use of AI. Research on AI policy frameworks emphasizes the importance of accuracy, clarity on sources of error, and communication of uncertainties in algorithms and data used to support decision-making. Transparent policies enable external examination and independent review of algorithms to ensure that AI systems operate as intended, free from harmful biases or errors. These policies highlight the importance of explainability in algorithmic decisions which can be facilitated by the use of explainable AI (XAI) tools that help visualize AI processes, interpret model predictions, and explain the relevance of input data in generating outputs.[78]

Professional development (PD) for faculty will be crucial to equip educators with the skills and knowledge needed to integrate AI into their teaching practices. Research on postsecondary PD underscores the importance of balancing academic freedom with the need for continuous professional growth. Studies highlight that top-down mandates often face resistance, suggesting that involving faculty in planning and implementation increases acceptance and participation.[79] A significant challenge is the integration of professional learning within the constraints of time and space, as well as the cost of such initiatives.[80] Innovative approaches that emphasize both collaborative learning and individualized training are recommended to improve trust and knowledge-sharing among faculty and to develop a learning environment that supports

---

[77] Deacon, B., Laufer, M., & Schäfer, L. O. (2022). Infusing educational technologies in the heart of the university—A systematic literature review from an organisational perspective. *British Journal of Educational Technology, 53*(6), 1932-1951.

[78] Nagy, M., & Molontay, R. (2024). Interpretable dropout prediction: Towards XAI-based personalized intervention. *International Journal of Artificial Intelligence in Education, 34*(1), 274–300.

[79] Golhasany, H., & Harvey, B. (2022). Academic freedom, the impact agenda, and pressures to publish: Understanding the driving forces in higher education. *SN Social Sciences, 2*(163).

[80] Fabriz, S., Hansen, M., Heckmann, C., Mordel, J., Mendzheritskaya, J., Stehle, S., … Horz, H. (2020). How a professional development programme for university teachers impacts their teaching-related self-efficacy, self-concept, and subjective knowledge. *Higher Education Research & Development, 40*(4), 738–752.

experimentation.[81] Knowledgeable mentors can play a crucial role in guiding faculty through these learning processes, ensuring that PD initiatives are both responsive and impactful.[82] Partnerships within institutions are equally important. A case study at Norfolk State University (NSU) demonstrates how collaboration between academic and student affairs offices can enhance student success, retention, and degree attainment, especially in under-resourced settings like HBCUs. Collaboration at NSU established a campus culture of innovation and cooperation that was supported by leadership. Efficient use of limited resources spurred innovation, allowing faculty and staff to discover and utilize hidden assets. Regular assessments provided insights into strengths and weaknesses, guiding improvements in collaborative efforts. These structural and cultural factors, along with a strong institutional mission, motivated and sustained collaborative efforts that led to positive student outcomes.[83]

Finally, system interoperability, data portability, and secure data sharing practices are key components in the successful application of AI in educational settings. For example, a learning management system (LMS) can seamlessly interact with AI tools to personalize learning experiences by accessing and analyzing student data such as grades, learning patterns, and participation levels.[84] Seamless data exchange allows students to receive customized learning experiences and provides educators with insights to improve instructional strategies. Data sharing practices should prioritize student privacy and the protection of sensitive information.

[81] Pischetola, M., Møller, J. K., & Malmborg, L. (2023). Enhancing teacher collaboration in higher education: The potential of activity-oriented design for professional development. *Education and Information Technologies, 28*(6), 7571–7600.

[82] Barry, W. (2022). The role of 'knowledgeable others' in supporting academics' professional learning: implications for academic development. *Perspectives: Policy and Practice in Higher Education, 27*(1), 16–25.

[83] Commodore, F., Gasman, M., Conrad, C., & Nguyen, T.-H. (2018). Coming together: A case study of collaboration between student affairs and faculty at Norfolk State University. *Frontiers in Education, 3.*

[84] Pesovski, I., Santos, R., Henriques, R., & Trajkovik, V. (2024). Generative AI for customizable learning experiences. *Sustainability, 16*, 3034.

Avoiding the Discriminatory Use of Artificial Intelligence

The U.S. Department of Education's (Department's) Office for Civil Rights (OCR) provides this resource to assist school communities with ensuring that artificial intelligence (AI) is used in a nondiscriminatory manner in the nation's elementary and secondary schools and institutions of higher education consistent with federal civil rights laws.[i] Within this resource, AI means a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments.[ii] AI technologies have the potential to enhance opportunities and increase educational equity for all students. At the same time, the growing use of AI in schools, including for instructional and school safety purposes, and AI's ability to operate on a mass scale can create or contribute to discrimination.

The Department's Office of Educational Technology (OET) has released several resources which provide helpful information about the use of artificial intelligence in education settings that are available on OET's website.

OCR enforces several federal civil rights laws that prohibit discrimination in programs or activities that receive federal financial assistance from the Department, including Title VI of the Civil Rights Act of 1964 (Title VI), which prohibits discrimination on the basis of race, color, or national origin; Title IX of the Education Amendments of 1972 (Title IX), which prohibits discrimination on the basis of sex; and Section 504 of the Rehabilitation Act of 1973 (Section 504), which prohibits discrimination on the basis of disability.[iii] OCR shares in the enforcement of Title II of the Americans with Disabilities Act of 1990 (Title II) with the U.S. Department of Justice. Title II prohibits discrimination on the basis of disability by state and local governments, regardless of whether or not those entities receive federal financial assistance.[iv] Each of the statutes OCR enforces and their implementing regulations also prohibits retaliation.[v]

The nondiscrimination provisions of these federal civil rights laws apply to discrimination resulting from the use of AI. This resource provides information regarding federal civil rights laws in OCR's jurisdiction and includes examples of types of incidents that could, depending on the facts and circumstances, present OCR with sufficient reason to open an investigation. Many of the examples included in this resource highlight conduct of schools and colleges that could also constitute discrimination without the use of AI technologies, but which may be compounded by its use. These examples are illustrative, non-exhaustive, and do not dictate the outcome of any particular matter OCR may investigate; rather, in each case, OCR engages in an individualized analysis of the particular facts at issue. OCR will open a case for investigation if the allegation(s) could state a violation of one of the laws or regulations that OCR enforces; this is an initial determination prior to investigation and not a finding or conclusion regarding

the allegations. If OCR opens an investigation, any determination about whether discrimination has occurred or is occurring would depend on particular facts and circumstances.

The contents of this guidance do not have the force and effect of law and do not bind the public or create new legal standards. This document is designed to provide clarity to the public regarding existing legal requirements under the civil rights laws that OCR enforces.

Anyone who believes that they or someone else has been discriminated against by a school or other recipient based on race, color, national origin, sex, disability, or age can file a complaint of discrimination with OCR using OCR's Electronic Complaint Assessment System.

Race, Color, or National Origin Discrimination:

Title VI states "no person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance."[vi]

Examples 1 through 7 below include the kinds of incidents that could, depending upon facts and circumstances, constitute race, color, or national origin discrimination under Title VI.

Title VI's protection extends to students with a shared ancestry or ethnic characteristics, including students from different countries or regions in the world. Title VI's protection also extends to students who are English Learners (ELs) and parents and guardians who have limited English proficiency (LEP). Specifically, Title VI requires elementary and secondary schools to provide appropriate language assistance services to students who are ELs that enable them to become proficient in English and participate equally and meaningfully in the standard instructional program within a reasonable period of time.[vii] Additionally, appropriate language assistance services must be educationally sound in theory and effective in practice.[viii]

OCR would likely have reason to open an investigation if a person filed a complaint based on the facts included in Examples 1 and 2 below, which involve students who are ELs:

> Example 1: A high school English teacher assigns the class a book report. After the students turn in their reports, the teacher grades them but decides to also run the reports through a free online service that inaccurately claims it can spot use of generative AI and prevent cheating and plagiarism. Unbeknownst to the teacher, the service has a low error rate for evaluating essays written by native English speakers for use of generative AI but has a high error rate when evaluating essays written by non-native English speakers. The detector flags as AI-generated the essays written by the class's only two students who are ELs, though they did not use AI or plagiarize. As a result, the teacher gives them both a failing grade and writes them up for academic dishonesty. A parent of one of the students appeals the student's grade and academic dishonesty charge. The principal tells the parent that she believes the AI-detection checker over the student's objections and that the school will continue to use the checker to ensure all papers are free from plagiarism. *OCR would likely have reason to open an investigation based on*

this complaint. Based on the facts, as alleged, the students who are ELs may not be able to participate equally and meaningfully in the standard instructional program of the school.

Example 2: An elementary school's supplemental English language development program for students who are ELs relies almost exclusively on a computer-based program that markets itself as a personalized learning program. This program will supplement instruction for students who are ELs in listening, speaking, reading, and writing. Twice a week, students who are ELs use this computer program for thirty minutes while non-EL students receive direct literacy intervention instruction from a classroom teacher. While the teachers are trained to assist if a student who is an EL has technical difficulty (e.g., login issues), they are not trained in providing English language development instruction and are unprepared to do so. A teacher files a complaint with OCR alleging that none of the students who are ELs are improving in their English language skills after two years in the EL program. *OCR would likely have reason to open an investigation based on this complaint.* Based on the facts, as alleged, the students who are ELs may not be able to participate equally and meaningfully in the standard instructional program of the school.

For more information regarding students who are ELs, please see:

- English Learner Students and Limited English Proficient Parents Dear Colleague Letter
- Ensuring English Learner Students Can Participate Meaningfully and Equally in Educational Programs Fact Sheet
- Equal Educational Opportunities for English Learners Website

Title VI requires that elementary and secondary schools ensure meaningful communication with parents and guardians who have LEP in a language they can understand. Elementary and secondary schools may take steps to provide effective language assistance to parents and guardians who have LEP by providing accurate written translations or oral interpretation.[ix]

OCR would likely have reason to open an investigation if a person filed a complaint based on the facts included in Example 3 below, which involves parents who have LEP:

Example 3: Teachers at a middle school often rely on non-contracted, online third-party applications and websites to translate and interpret for parents who have LEP. Some of these applications leverage AI to translate or interpret. A parent complains to the school's principal that they have been unable to communicate with their child's teacher regarding the child's academic and behavioral progress in class due to incoherent translations. The school does not investigate or attempt to resolve the parent's concern. *OCR would likely have reason to open an investigation based on this complaint*. Based on the facts, as alleged, the school may not be ensuring meaningful communication with parents who have LEP in a language the parents can understand.

For more information regarding parents and guardians who have LEP, please see:

- English Learner Students and Limited English Proficient Parents Dear Colleague Letter
- Information for Limited English Proficient Parents and Guardians and for Schools and School Districts that Communicate with Them Fact Sheet
- Equal Educational Opportunities for English Learners Website

Harassment creates a hostile environment under Title VI if school staff, a student, or another person engages in unwelcome conduct based on race, color, or national origin that, based on the totality of the circumstances, is subjectively and objectively offensive and is so severe or pervasive that it limits or denies a person's ability to participate in or benefit from the recipient's education program or activity. OCR could find a Title VI violation if: (1) a hostile environment based on race, color, or national origin existed; (2) the school had actual or constructive notice (in other words, the school knew or should have known) of the hostile environment; and (3) the school failed to take prompt and effective steps reasonably calculated to (i) end the harassment, (ii) eliminate any hostile environment and its effects, and (iii) prevent the harassment from recurring.[x]

OCR would likely have reason to open an investigation if a person filed a complaint based on the facts included in Example 4 below, which involves harassment under Title VI:

> Example 4: A school district purchases facial recognition technology from a third-party vendor that markets its product as essential for school safety. The school district is unaware that the facial recognition technology it purchased has difficulty accurately recognizing the faces of students of color and consistently misidentifies Black individuals. A school in the district compiles a persons of interest list that includes individuals arrested for drug usage in the area and individuals who have been issued trespass warnings, including previously expelled students. If the facial recognition system detects that a person of interest is on campus, it sends an alert to specified school personnel. The system mistakenly flags several Black students multiple times throughout the school year, causing each student to be stopped and questioned by the school's resource officer and administrators. The students complain to the principal that they are being falsely flagged and questioned. The students flagged are also pulled out of class on multiple occasions to verify that they are not on the persons of interest list. The school does not take any actions in response to the students' complaints or to resolve the students' missed class time. Other students begin calling Black students "criminals" based on their continued interactions with the resource officer. *OCR would likely have reason to open an investigation based on this complaint*. Based on the facts, as alleged, the students may have experienced a hostile environment due to the multiple false flags, being pulled out of class and questioned, and the allegations regarding peer harassment.

For more information, please see:

- Harassment Website
- Harassment based on Race, Color, or National Origin on School Campuses Fact Sheet
- Race and School Programming Dear Colleague Letter

Under Title VI, discrimination can occur when a school implements or enforces a policy or practice in a manner that treats students differently based on race, color, or national origin. OCR frequently uses the following three step test to determine whether a school treated students differently based on race, color, or national origin in violation of Title VI:

1. Did the school treat a student or group of students of a particular race, color, or national origin differently from a similarly situated student or group of students of another race, color, or national origin? If no, then OCR would not find sufficient evidence to determine that the school has engaged in different treatment under this framework. If yes, then move to step two:
2. Can the school articulate a legitimate, nondiscriminatory reason for the different treatment? If no, OCR could find that the school has discriminated on the basis of race, color, or national origin. If yes, then move to step three:
3. Is the articulated reason for the different treatment a pretext for discrimination (i.e., not the true reason for the school's actions)? If yes, OCR could find that the school has engaged in discrimination based on race, color, or national origin.

OCR would likely have reason to open an investigation if a person filed a complaint based on the facts included in Examples 5 and 6 below, which involve discrimination in student discipline:

> Example 5: A school implements new risk assessment, AI-enabled software to determine appropriate discipline measures for students. The software gives students a risk score that estimates the likelihood of them committing a future severe infraction based on historic school discipline data. Significant disparities by race have persisted in the school's application of student discipline, and Black students are disciplined more frequently and more harshly than other similarly situated students of another race. As a result, the historic school discipline data that the software relies on reflects the school's discriminatory disciplinary practices. Whenever a student receives a disciplinary referral, the school uses the software to calculate the student's chance of reoffending and heavily relies on the risk score in determining the severity of punishment (for example, higher scores lead to more significant discipline). The AI software does not directly use the student's race as an input in calculating the risk score. Parents complain to the school after they notice that the software tends to give high scores to Black students and low scores to White students. The principal tells the parents they should trust the software because it does not have access to information about the student's race, so it is less biased to let a machine generate the risk score for discipline decisions than to let a school official do it. *OCR would likely have reason to open an investigation based on this complaint*. Based on the facts, as alleged, the school may be treating Black students differently than similarly situated students of another race.
>
> Example 6: A high school purchases AI-enabled software that uses predictive analytics to identify students with a higher likelihood of dropping out of school in order to monitor their progress. The software uses a number of inputs to make its determination including whether the student has a two-parent household, whether the student has a low attendance record, whether the student has a disciplinary record, and if the student is a member of a racial or ethnic minority group. As a result, the software flags a large portion of the school's Black and Hispanic/Latino student population. The software places red flags next to the identified students' names on classroom rosters. In preparation for college visitation week, the school's

assistant principal uses the outputs from the program to exclude red-flagged students based on his belief that they are unlikely to attend college and that it would be a waste of time. A Latino parent files a complaint against the school after their child is automatically excluded from college visitation events. *OCR would likely have reason to open an investigation based on this complaint*. Based on the facts, as alleged, the school may be treating Black and Hispanic/Latino students differently than similarly situated students of other races.

For more information, please see:

- Discriminatory Discipline Website
- Confronting Racial Discrimination in Student Discipline Resource

Sex Discrimination:

OCR also enforces Title IX of the Education Amendments of 1972 (Title IX), which prohibits discrimination on the basis of sex in federal funding recipients' education programs and activities. The Department's Title IX regulations were amended in 2020 (2020 Notice of Final Rule; 2020 Title IX regulations) and again in 2024 (2024 Notice of Final Rule; 2024 Title IX regulations). As of September 13, 2024, pursuant to Federal court orders, the Department is enjoined from enforcing the 2024 Title IX regulations in certain states and schools. Pending further court orders, the Department's Title IX regulations, as amended in 2020, remain in effect in those states and schools.[xi]

For more information, please see:

- Sex Discrimination: Overview
- Resources for LGBTQI+ Students
- Frequently Asked Questions About Sex Discrimination

Examples 7 through 12 include the kinds of incidents that could, depending on facts and circumstances, constitute sex discrimination under Title IX.

Under the 2024 Title IX regulations, sex-based harassment creates a hostile environment if school staff, a student, or another person engages in unwelcome sex-based conduct that, based on the totality of the circumstances, is subjectively and objectively offensive and is so severe or pervasive that it limits or denies a person's ability to participate in or benefit from the recipient's education program or activity.[xii]

Under the 2020 Title IX regulations, sexual harassment creates a hostile environment if school staff, a student, or another person engages in unwelcome sex-based conduct that is so severe, pervasive, and objectively offensive that it effectively denies a person equal access to the recipient's education program or activity.

OCR would likely have reason to open an investigation of the allegations described in Examples 7 and 8 under the 2020 Title IX regulations or 2024 Title IX regulations.

OCR would likely have reason to open an investigation if a person filed a complaint based on the facts included in Examples 7 and 8 below, which involve harassment under Title IX:

> Example 7: A student uses a generative AI application to send repeated and continuous sexually explicit messages to a classmate's school email. That classmate's parents report the issue to administrators, asking the school to take actions to stop the messages. The principal states that because she does not have access to the tool the student is using to spam their classmate's email, she cannot investigate further or prevent the messages from being sent. *OCR would have reason to open an investigation based on this complaint*. Based on the facts, as alleged, the student receiving the sexually explicit messages may have experienced prohibited harassment about which the school knew and failed to appropriately respond.
>
> Example 8: A high school student discovers that a public anonymous social media account is posting images of a female student and other classmates that depict photographs of the students' faces with AI-created nude bodies (sometimes referred to as "deepfake nudes") and include sexually explicit comments and tags. Each new post by the account becomes a popular topic of discussion amongst students at the school. Another student brags to friends at the school about creating the account and then is reported to administrators. A third student reports the social media account to her parents who in turn file a complaint with the school. The school tells the parents that they are aware of the account due to multiple reports and have reported it to the police but that they cannot do anything further to ameliorate the situation until the police complete their investigation. *OCR would have reason to open an investigation based on this complaint*. Based on the facts, as alleged, the students may have experienced prohibited harassment about which the school knew and failed to appropriately respond.

In addition to allegations of sex-based harassment that creates a hostile environment, under Title IX, OCR can investigate allegations that a school is implementing or enforcing a policy or practice in a manner that treats students differently based on sex. When determining whether a school treated students differently based on sex in violation of Title IX, OCR frequently uses a three-step test similar to the one described above in Examples 5 and 6.

OCR would likely have reason to open an investigation if a person filed a complaint based on the facts included in Examples 9 through 11 below, which involve a school implementing or enforcing a policy or practice in a manner that treats students differently based on sex:

> Example 9: A school uses AI software to make students' academic schedules. The software relies on historic data and students' demographics to determine course enrollments. Historically, male students were twice as likely to take computer science as compared to female students. The software creates schedules that place only one female student in the school's computer science class, which has thirty-five students. Other female students who want to take the computer science class ask the principal to re-do their schedules so they can take the course, but the principal declines, saying that it would be too time consuming to re-do their schedules. *OCR would have reason to open an investigation based on this complaint.* Based on the facts, as alleged, the female students may have been treated differently than male students, who are otherwise similarly situated.

Example 10: A public college utilizes an AI-enabled software that uses predictive analytics to identify students who will be most successful in its engineering program, and then assigns each applicant a score. The score is weighted heavily in deciding whether to admit students to the program, with lower weights being given to a student essay and answers provided on an application, which, among other things, includes the students' demographic information. The software uses data from past engineering applicants' acceptances and their demographics as an input for its analysis. Nearly 90% of past students are men. Though approximately equal numbers of male and female students apply, approximately 80% of students admitted to the program are male students. Several female applicants complain to administrators. The college dean tells them that he has full faith in the software and believes it is more than fair that 20% of the students admitted are female, since that is twice the percentage of female engineers in the state. *OCR would have reason to open an investigation based on this complaint*. Based on the facts, as alleged, the female students may have been treated differently than similarly situated male students.

Example 11: A school uses facial recognition technology to check students into the school. The technology repeatedly flags students as a security risk if they do not conform to the technology's assumptions as to what girls and boys should look like, based on the sex specified in their school records. Students who are falsely flagged must wait for a school administrator to override the software, resulting in embarrassment and missed class time. The principal is aware of the problem but decides to keep using the technology due to perceived security benefits. *OCR would have reason to open an investigation based on this complaint*. Based on the facts, as alleged, the students may have been treated worse than similarly situated students.

Under Title IX, a recipient that operates or sponsors interscholastic athletics must provide equal athletic opportunity, regardless of sex.[xiii]

OCR would likely have reason to open an investigation if a person filed a complaint based on the facts included in Example 12 below, which potentially involves an unequal provision of athletic opportunities:

Example 12: A high school begins utilizing an AI tool to assist with the scheduling for athletic practices and games. The AI tool uses historical data from previous school year schedules to generate proposed schedules for boys' and girls' teams. A female basketball player complains to the school's athletic director that the girls' basketball team always plays on Tuesday or Wednesday night while the boys' basketball team consistently plays on Friday nights and Saturdays and therefore does not need to prepare for class the days after their games. The student also shares that the boys' basketball team receives all of the earlier practice times, and the girls' basketball team ends up practicing at 8:00 PM on weeknights. Although the athletic director retains oversight for scheduling, she tells the student that the AI tool is responsible for scheduling this year and that she does not know how to change it. She does not respond further to the student's complaint. *OCR would have reason to open an investigation based on this complaint.* Based on the facts, as alleged, the school may not be providing equal athletic opportunity to members of the girls' basketball team.

Disability Discrimination:

Section 504 of the Rehabilitation Act of 1973 (Section 504) states that no otherwise qualified individual with a disability shall, solely by reason of their disability, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving federal financial assistance.[xiv]

For more information, please see:

- Disability Discrimination Issues Website
- Students with Disabilities Preparing for Postsecondary Education Resource
- Frequently Asked Questions on Effective Communication for Students with Hearing, Vision, or Speech Disabilities in Public Elementary and Secondary Schools

Examples 13 through 18 include the kinds of incidents that could, depending upon facts and circumstances, constitute disability discrimination under Section 504.

Section 504 requires colleges, universities, and other postsecondary institutions to provide equal opportunities to people with disabilities in all their operations, including equal opportunities to participate in and benefit from online programs and activities. Under Section 504, these recipients of Federal financial assistance must provide such academic adjustments, which can include reasonable modifications or auxiliary aids and services, as may be necessary to ensure that they do not discriminate on the basis of disability.

OCR would likely have reason to open an investigation if a person filed a complaint based on the facts included in Example 13 below, which involves academic adjustments for a postsecondary student:

> Example 13: An AI test proctoring software uses facial recognition technology and eye movement tracking to monitor students for behavior that indicates they might be cheating during exams. A student with a disability receives a failing grade for an exam after the software flags her behavior as suspicious and her professor accuses her of cheating. The student appeals the grade because her vision impairment causes eye movements that the software falsely flagged as suspicious. The student also requests that she receive an academic adjustment so that she does not have to take tests using this particular proctoring software. The university threatens to expel the student if she is flagged for the same behavior again and does not respond to the student's request. *OCR would have reason to open an investigation based on this complaint*. Based on the facts, as alleged, the student may not have been provided with necessary academic adjustments.

OCR would likely have reason to open an investigation if a person filed a complaint based on the facts included in Example 14 below, which involves academic adjustments in the form of auxiliary aids and services for a postsecondary student:

> Example 14: A deaf student enrolls in a university's engineering program. The university provides the student with an AI-aided closed circuit captioning transcription application for class lectures. The student repeatedly informs the university with specific detail that the AI service provided to transcribe the lectures does not accurately capture the advanced engineering terminology essential to her education program. The university does not provide the student

alternative auxiliary aids and services to access the class lectures. *OCR would have reason to open an investigation based on this complaint.* Based on the facts, as alleged, the student may not have been provided with necessary academic adjustments.

For more information, please see:

- Disability Discrimination Issues Website
- Parent and Educator Resource Guide to Section 504 in Public Elementary and Secondary Schools
- Protecting Students With Disabilities FAQ

Section 504 requires public elementary and secondary schools to provide students with disabilities with a free appropriate public education (FAPE). Under Section 504, FAPE is the provision of regular or special education and related aids and services that: (1) are designed to meet the individual educational needs of students with disabilities as adequately as the needs of students without disabilities are met; and (2) are based on adherence to procedures governing educational setting, evaluation and placement, and procedural safeguards.[xv] Parents and guardians have the right to challenge FAPE decisions using Section 504's procedural safeguards.[xvi]

OCR would likely have reason to open an investigation if a person filed a complaint based on the facts included in Examples 15 and 16 below, which involve the provision of FAPE to elementary and secondary students:

Example 15: A school staff member utilizes an AI-driven adaptive assessment to determine admission to the school's gifted program. The assessment's questions get more difficult if a student answers quickly and correctly, and students' scores on the tests are determined, in part, by how quickly students answer difficult questions. During the assessment, a student with attention-deficit/hyperactivity disorder (ADHD) is distracted and takes longer to answer questions, causing the assessment to produce easier questions, and thus reflects his disability rather than his actual aptitude or achievement. Although the student has an Individualized Education Program (IEP) with test taking accommodations including increased time, the teacher does not believe the IEP applies to the gifted admissions assessment. The student is deemed ineligible for the gifted program. The student's parent files a complaint alleging that the student's final score is lower than it would have been if the teacher had not used the AI-driven adaptive assessment and if the student had been provided with his test taking accommodations. *OCR would have reason to open this complaint for investigation.* Based on the facts, as alleged, the student may have experienced disability discrimination in the admissions process and/or not have been provided with FAPE.

Example 16: A school district allows schools in the district to use a generative AI tool to write Section 504 Plans for students with disabilities. The school district does not have any policies regarding how to use the tool or how to ensure that the group of knowledgeable people responsible for evaluating a student review what the AI produces to determine whether it meets the individual needs of each student. One school begins using the tool to create Section 504 Plans for all students with diabetes. School staff do not review or modify the generated Section 504 Plans and begin implementing them, and they inform parents that they believe AI tools make more effective choices than people. A local group of parents of students with diabetes at that school files a complaint with the school district stating that their students' Section 504

Plans' provisions look almost identical and, in some cases, do not match the specific needs of their children. The school district states that they defer to the school's decision on how to utilize AI tools and does not investigate further. *OCR would have reason to open this complaint for investigation.* Based on the facts, as alleged, students may not have been provided with FAPE because their 504 plans may not have been designed to meet their individual educational needs.

For more information, please see:

- Supporting Students with Disabilities and Avoiding the Discriminatory Use of Student Discipline under Section 504 of the Rehabilitation Act of 1973 Dear Colleague Letter
- Supporting Students with Disabilities and Avoiding the Discriminatory Use of Student Discipline Under Section 504 of the Rehabilitation Act of 1973 Fact Sheet
- Parent and Educator Resource Guide to Section 504 in Public Elementary and Secondary Schools

Section 504 requires schools to provide the services, supports, interventions, strategies, and modifications to school or district policies for students with disabilities to address any disability-based behavior, including behavior that could lead to discipline, and failure to do so may constitute a denial of FAPE. OCR's continued enforcement experience reflects that many students with disabilities also face discipline because they are not receiving those supports, services, interventions, strategies, and modifications to school or district policies that they need to manage their disability-based behavior. In addition to that discrimination, many students with disabilities are also subjected to discrimination based on their disability when being disciplined, such as when students with disabilities are unnecessarily disciplined more severely than students without disabilities for the same or similar behavior. When schools do discipline students with disabilities, they must do so in a nondiscriminatory manner.

OCR would likely have reason to open an investigation if a person filed a complaint based on the facts included in Example 17 below, which involves discipline-related issues for a student with a disability:

Example 17: A middle school uses content moderation software to alert the school if any language that violates the student code of conduct is used on school-issued devices. A student with obsessive-compulsive disorder has a compulsion to say certain swear words if certain inciting incidents occur, and using some of the words would otherwise violate the student code of conduct. The student's Section 504 Plan addresses the techniques the student and staff are using to work on this compulsion as well as what happens when the student uses the identified words. The student uses his school-issued device to chat with a peer and uses some of the words identified in his Section 504 Plan that would otherwise violate the student code of conduct. The content moderation software flags this language and alerts school officials that problematic language has been detected on his school-issued device. Based only on the flag from the content moderation software, the principal immediately punishes the student without consulting or following the procedures in the 504 Plan. The parent of the student files a complaint against the school stating that the school's disciplinary response denied the student FAPE. *OCR would have reason to open this complaint for investigation.* Based on the facts, as alleged, the school did not follow the student's 504 Plan and, and as a result, the student may not have been provided with FAPE.

Harassment creates a hostile environment under Section 504 if school staff, a student, or another person engages in unwelcome conduct based on disability that, based on the totality of the circumstances, is subjectively and objectively offensive and is so severe or pervasive that it limits or denies a person's ability to participate in or benefit from the recipient's education program or activity. OCR could find a Section 504 violation in its enforcement work if: (1) a hostile environment based on disability existed; (2) the school had actual or constructive notice (in other words, the school knew or should have known) of the hostile environment; and (3) the school failed to take prompt and effective steps reasonably calculated to (i) end the harassment; (ii) eliminate any hostile environment and its effects; and (iii) prevent the harassment from recurring.

OCR would likely have reason to open an investigation if a person filed a complaint based on the facts included in Example 18 below, which involves harassment under Section 504:

> Example 18: An elementary school teacher uses an AI enabled application, which monitors noise and provides feedback to assist her in managing classroom noise. The application uses the class computer's built-in microphone to detect when students' voices are raised, and it displays a color meter ranging from green (quiet classroom) to red (loud classroom). The application tracks patterns for class noise and predicts the times of day that the teacher is likely to have more difficulty managing classroom noise. During the times of day that the application predicts will be the loudest, the teacher keeps the color meter projected for all the students to see and offers a pizza party if they only have a few instances where the meter reaches red. A student who is hard of hearing reports that they are being bullied by a few classmates because they believe the student's speaking voice consistently causes the meter to be red and the class does not receive a pizza party as a result. The school does not respond to the student's report, and the teacher tells the student that they need to learn to speak more quietly. *OCR would have reason to open this complaint for investigation.* Based on the facts, as alleged, the student may have experienced prohibited harassment about which the school knew and failed to appropriately respond.

Multiple Bases Discrimination:

Examples 19 through 21 include the kinds of incidents that could, depending upon facts and circumstances, constitute discrimination pertaining to multiple bases discussed above in this resource.

> Example 19: A school district purchases an AI-driven application to streamline the universal screening process for speech and language disorders. The school district decides to only utilize the AI-driven application and not employ or seek opinions from Speech Language Pathologists, or other appropriate professionals. The application falsely flags students who are ELs as students with a speech disorder. The district refuses to evaluate students for speech and language disabilities, if a student was not identified through the universal screening, unless the student's parent obtains and submits a private diagnosis. The application also misses several students who have speech and language disorders. Although parents complain to the school about the inaccuracy of the AI-driven determinations, the school does not change its process. *OCR would have reason to open this complaint for investigation.* Based on the facts, as alleged, students who are ELs may not be able to equally and meaningfully participate in the standard instructional program. OCR would also investigate whether the school district's use of the AI-

driven application is erroneously impacting its obligations to identify and locate qualified students with disabilities, to evaluate them, and provide FAPE.

Example 20: A school district starts using an internal software program to generate student Individualized Education Programs (IEPs). The software is trained on past IEPs to recommend appropriate placements for current students. The software inputs include all available demographic information about students, including race. Historically, more Black students with disabilities in the district had IEPs that included more hours of special education instruction in a separate setting and educational placements that were more restrictive than other students. Most of the IEPs for Black students that are generated by the software recommend more special education services in separate settings and would result in placements in restrictive educational environments, but the IEPs that the software generates for white students with similar disability-related needs recommends more integrated instruction and would result placements in less restrictive educational environments. A special education teacher complains to the principal that the software is drafting inappropriate IEPs for many of the Black students with disabilities, but the principal says the software is just doing its job and they do not have the resources to review every IEP to ensure that it is appropriate for each student. *OCR would have reason to open this complaint for investigation.* Based on the facts, as alleged, the school may be treating Black students differently than similarly situated white students, and students with disabilities may not be receiving FAPE.

Example 21: A high school purchases AI software that electronically tracks how often students sign out for hall passes to estimate students' mental well-being. The software flags students who sign out an electronic hall pass more than three times a day as a factor related to mental well-being. If the algorithm determines that a student has a low mental well-being score, the student is pulled out of class to have mandatory meetings with the school counselor. During the school's first week using the software, the algorithm gives low mental well-being scores to students who are menstruating, pregnant students, and students with gastrological disabilities, since those students needed to use the bathroom more than their peers. The students tell the principal that they would like to return to their classes and that they do not believe meetings with the counselor are needed or helpful, but the principal indicates that whether the meetings are needed or not, he cannot ignore a potential issue with a student's mental well-being flagged by the software. As a result, the students miss substantial learning time because of the mandatory meetings with the school counselor. *OCR would have reason to open this complaint for investigation.* Based on the facts, as alleged, the school may be treating students who are menstruating or pregnant differently than students who are not menstruating or pregnant. The school may also be failing to implement provisions or discouraging students from taking advantage of necessary provisions, such as unlimited bathroom breaks, in the Section 504 Plan of a student with a gastrological disability.

For more information regarding students who are ELs, pregnancy and related conditions, and protections for students with GER or GERD:

- English Learner Students and Limited English Proficient Parents Dear Colleague Letter
- Ensuring English Learner Students Can Participate Meaningfully and Equally in Educational Programs Fact Sheet
- Equal Access to Elementary and Secondary Education for Students Who Are English Learners with Disabilities Fact Sheet
- Equal Educational Opportunities for English Learners Website
- Discrimination Based on Pregnancy and Related Conditions
- Section 504 Protections for Students with GER or GERD

Schools' use of AI and AI-driven technologies may also raise concerns pertaining to the Family Educational Rights and Privacy Act (FERPA), a statute not enforced by OCR. For more information on FERPA, please see the Department's FERPA website.

Anyone who believes that they or someone else has been discriminated against by a school or other recipient based on race, color, national origin, sex, disability, or age can file a complaint of discrimination with OCR using OCR's Electronic Complaint Assessment System. Complaint forms and other resources, available in languages other than English, are available on OCR's Resources Available in Other Languages website.

All of the federal civil rights laws enforced by OCR prohibit retaliation. In general, retaliation occurs when a school intimidates, threatens, coerces, or discriminates against an individual for one of two reasons: (1) Because the individual has made a complaint or testified, assisted, or participated in any manner in an OCR investigation or proceeding, or (2) For the purpose of interfering with any right or privilege under federal civil rights law enforced by OCR.

For more information on the prohibitions against retaliation under federal civil rights laws, see OCR's website on Retaliation.

To request technical assistance—including webinars, trainings, or presentations at conferences or other local events—concerning the legal obligations for schools and colleges under federal civil rights laws, you can contact OCR at 800-421-3481 (TDD: 800-877-8339), at OCR@ed.gov, or through the contact information for OCR's regional offices available on the OCR Addresses Webpage. School boards and other related governmental agencies may also contact the regional Equity Assistance Centers, which are funded by the Department and provide technical assistance and training, in the nondiscrimination assistance areas of race, sex, national origin, and religion to promote equitable education opportunities.[xvii]

To request language access services or resources, which may include oral technical assistance or written translation of Department information, free of charge, contact OCR@ed.gov. If you need more information about interpretation or translation services, call 1-800-USA-LEARN (1-800-872-5327) (TTY: 1-800-437-0833). To request documents in alternate formats such as Braille or large print, contact the Department at 202-260-0852 or om_eeos@ed.gov.

[i] Within this resource, "school(s)" or "college(s)" is used generally to refer to elementary, secondary, and postsecondary educational institutions that are recipients of federal financial assistance from the Department.

[ii] Artificial intelligence systems use machine- and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information or action. *See* E.O. 14110, Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 88 Fed. Reg. 75191-75226 (October 30, 2023); 15 U.S.C. 9401(3).

[iii] 42 U.S.C. § 2000d, *et seq.;* 20 U.S.C. § 1681 *et seq.;* 29 U.S.C. § 794. OCR also enforces the Age Discrimination Act of 1975, 42 U.S.C. § 6101 *et seq*., and the Boy Scouts of America Equal Access Act, 20 U.S.C. § 7905. This letter does not specifically address those statutes.

[iv] Title II of the Americans with Disabilities Act of 1990 (ADA), 42 U.S.C. §§ 12131-12134, and DOJ's implementing regulation for Title II, 28 C.F.R. pt. 35, prohibit disability discrimination by State and local governments, regardless of whether they receive Federal funds. 42 U.S.C. § 12131 *et seq*. OCR and the U.S. Department of Justice share in the enforcement of Title II in public elementary and secondary education systems and institutions, public institutions of higher education, vocational education (other than schools of medicine, dentistry, nursing, and other health-related schools), and public libraries. 28 C.F.R. §§ 35.172–35.174, 35.190(b)(2). When discussing uses of AI that could discriminate against people with disabilities, this document focuses on Section 504. However, entities covered by both Section 504 and the ADA should consider their disability nondiscrimination obligations under both laws. Section 504 and Title II are generally interpreted consistently, though DOJ's Title II regulation sometimes includes more specificity. For more information about Title II of the ADA, see DOJ's ADA.gov website.

[v] *See* 34 C.F.R. § 100.7(e) (Title VI); 34 C.F.R. § 106.71 (Title IX) (incorporating 34 C.F.R. §100.7(e) by reference); 34 C.F.R. § 104.61 (Section 504) (incorporating 34 C.F.R. §100.7(e) by reference); and 34 C.F.R. §108.9 (Boy Scouts Act) (incorporating 34 C.F.R. §100.7(e) by reference). Title II and the Age Act have similar regulatory language. See 28 C.F.R. § 35.134 (Title II); and 34 C.F.R. § 110.34 (Age Act).

[vi] 42 U.S.C. § 2000d *et seq*.

[vii] *See Lau v. Nichols*, 414 U.S. 563 (1974); *Castañeda v. Pickard*, 648 F.2d 989 (5th Cir. 1981); 42 U.S.C. § 2000d to d-7 (prohibiting race, color, and national origin discrimination in any program or activity receiving federal financial assistance); U.S. Dep't of Educ., U.S. Dep't of Just., *Ensuring English Learner Students Can Participate Meaningfully and Equally in Educational Programs*, https://www2.ed.gov/about/offices/list/ocr/docs/dcl-factsheet-el-students-201501.pdf.

[viii] *Id.*

[ix] *Id*.

[x] *See* Racial Incidents and Harassment against Students at Educational Institutions: Investigative Guidance, 59 Fed. Reg. 11,448, 11,449 (Mar. 10, 1994); OCR's Guidance on Schools' Obligations to Protect Students from Student-on-Student Harassment on Basis of Sex; Race, Color and National Origin; and Disability at 2, 4, 6 (Oct. 2010); OCR's Dear Colleague Letter on Race and School Programming (Aug. 2023); OCR's Dear Colleague Letter on Protecting Students from Discrimination, such as Harassment Based on Race, Color, or National Origin, including Shared Ancestry or Ethnic Characteristics (May 2024); OCR's Fact Sheet on Harassment based on Race, Color, or National Origin on School Campuses (July 2024).

[xi] The Department amended its Title IX regulations effective August 1, 2024. 89 Fed. Red. 33474 (Apr. 29, 2024). The 2024 Title IX regulations apply to alleged sex discrimination that occurred on or after August 1, 2024. As of September 13, 2024, pursuant to Federal court orders, the Department is currently enjoined from enforcing the 2024 Title IX regulations in the states of Alabama, Alaska, Arkansas, Florida, Georgia, Idaho, Indiana, Iowa, Kansas, Kentucky, Louisiana, Mississippi, Missouri, Montana, Nebraska, North Dakota, Ohio, Oklahoma, South Carolina, South Dakota, Tennessee, Texas, Utah, Virginia, West Virginia, and Wyoming; the Department is also currently enjoined from enforcing the 2024 Title IX regulations at the schools on the list located at https://www.ed.gov/sites/ed/files/about/offices/list/ocr/docs/list-of-schools-enjoined-from-2024-t9-rule. Per court order, this list of schools may be supplemented in the future. Pending further court orders, the Department's Title IX regulations, as amended in 2020, remain in effect in those states and schools. Any updates on the status of the 2024 Title IX regulations will be posted at https://www.ed.gov/laws-and-policy/civil-rights-laws/title-ix-and-sex-discrimination/sex-discrimination-overview-of.

[xii] *See* 34 C.F.R. 106.2; *See also* 89 Fed. Reg. 33,474, 33,515 (April 29, 2024).

[xiii] 34 C.F.R. § 106.41(c).

[xiv] 29 U.S.C. § 794.

[xv] *See* 34 C.F.R. § 104.33. Under the Section 504 regulations, one way to meet Section 504 requirements for a free appropriate public education is to implement an individualized education program pursuant to the Individuals with Disabilities Education Act (IDEA). *See* 34 C.F.R. § 104.33(b)(2). OCR does not enforce IDEA, but rather enforces the Section 504 and Title II rights of IDEA students with disabilities. For more information on IDEA, please visit Individuals with Disabilities Education Act (IDEA).

[xvi] *See* 34 C.F.R. § 104.36.

[xvii] *See* 34 C.F.R. part 270; https://oese.ed.gov/offices/office-of-formula-grants/program-and-grantee-support-services/training-and-advisory-services-equity-assistance-centers/.